

OM Of: US-09-294-298-6 to: GenEmbl:\* out\_format : pfs

Date: Jan 19, 2001 1:07 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

#### Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool/US09294298/runat\_17012001\_125501\_27756/app\_query.fasta\_1.4359  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blos62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=200000000 -USER=US09294298 @CGN1\_1\_13912 -NCPU=6  
-ICPU=3 -LONGLOG -NO\_XLPXY -WAIT -THREADS=1

#### Search information block:

Query: US-09-294-298-6  
Query length: 1325  
Database: GenEmbl:\*  
Database sequences: 1118133  
Database length: -1736092196  
Search time (sec): 10440.680000

score_list:	Sequence	Strd Orig	Zscore	EScore	Len	Documentation
gb_ro:AF050183	+ 6950.00	2848.35	1.7e-150	4265	1	AF050183 Rattus norvegicus GTH
gb_ro:AF058790	+ 6912.50	2833.25	1.2e-149	4140	1	AF058790 Rattus norvegicus Syn
gb_ro:AF058789	+ 6858.00	2810.58	2.2e-148	4539	1	AF058789 Rattus norvegicus Syn
gb_ro:AF048976	+ 6292.50	2581.05	1.3e-135	4063	1	AF048976 Rattus norvegicus syn
gb_ro:AF016962	+ 6280.50	2575.28	2.7e-135	4801	1	AF016962 Rattus norvegicus mRNA
gb_htg20:AL161903	- 4060.50	1653.62	6.0e-84	150956	1	AL161903 Homo sapiens chromo
gb_pr7:HSJ570F3	- 4059.50	1655.68	4.6e-84	94770	1	AL050332 Human DNA sequence f
gb_htg20:AL161906	+ 3987.00	1623.11	3.0e-82	169323	1	AL161906 Homo sapiens chromo
gb_pr3:AF047711	+ 2584.50	1071.76	1.5e-51	4368	1	AF047711 Homo sapiens ngap mRNA
gb_htg7:AC019800	- 1257.50	518.21	1.0e-20	56366	1	AC019800 Drosophila melanogas
gb_htg4:AC012162	+ 1257.50	511.73	2.4e-20	191504	1	AC012162 Drosophila melanoga
gb_htg4:AC012161	+ 1257.50	511.43	2.5e-20	202741	1	AC012161 Drosophila melanoga
gb_in1:AE003506	+ 1257.50	509.33	3.3e-20	300994	1	AE003506 Drosophila melanoga
gb_in1:AB011280	+ 1173.50	498.06	1.4e-19	3984	1	AB011280 Caenorhabditis elega
gb_pr4:AK022662	+ 1088.50	466.19	8.3e-18	2385	1	AK022662 Homo sapiens cDNA FLJ
gb_htg6:AC016523	+ 1082.00	445.96	1.1e-16	65961	1	AC016523 Drosophila melanog
gb_pr4:AK024488	+ 1081.50	440.24	1.8e-17	4287	1	AK024488 Homo sapiens mRNA for
gb_ro:AF053938	+ 761.00	338.52	1.1e-10	829	1	AF053938 Rattus norvegicus p135
gb_htg22:AL365274	- 710.00	289.50	5.7e-08	172027	1	AL365274 Homo sapiens chromo
gb_in2:CEC07B5	+ 676.50	283.64	1.2e-07	39660	1	Z46266 Caenorhabditis elegans
gb_htg12:AC027051	+ 650.50	265.78	1.2e-06	156700	1	AC027051 Homo sapiens chromo
gb_htg3:AC011492	+ 613.00	250.49	8.5e-06	157803	1	AC011492 Homo sapiens chromo
gb_htg17:AC073765	+ 600.00	243.03	2.2e-05	237619	1	AC073765 Mus musculus clone
gb_in1:AB011279	+ 503.50	228.38	0.0001	2276	1	AB011279 Caenorhabditis elega
gb_htg3:AC011703	+ 502.00	205.25	0.0028	159777	1	AC011703 Drosophila melanoga
gb_pr1:AB011110	+ 488.50	217.53	0.0006	5579	1	AB011110 Homo sapiens mRNA for
gb_htg13:AC040900	+ 479.50	196.24	0.0090	155332	1	AC040900 Homo sapiens chromo
gb_htg20:AL161906	- 479.00	195.58	0.0098	169323	1	AL161906 Homo sapiens chromo
gb_ro:RATGAPX	+ 455.00	206.69	0.0024	3296	1	L13151 Rat GTPase-activating p
gb_pr8:HUMGAPB	+ 437.50	199.31	0.0061	3456	1	M23612 Human GTPase-activating
gb_pr8:HUMGAPA	+ 437.50	198.15	0.0070	4307	1	M23379 Human GTPase-activating
gb_p12:BGR272158	+ 433.50	198.19	0.0070	3144	1	AJ272158 Blumeria graminis f.
gb_om:BTGAPR	+ 433.50	196.32	0.0089	4473	1	X12602 Bovine mRNA for GTPase
gb_pat1:EI13125	+ 429.50	197.47	0.0077	2647	1	E13125 Bovine gene for R-Ras p
gb_om:BTU30857	+ 429.50	197.31	0.0078	2730	1	U30857 Bos taurus R-Ras GTPase
gb_htg12:AC027051	- 417.00	170.76	0.2355	156700	1	AC027051 Homo sapiens chromo
gb_in3:DME012609	+ 416.00	190.49	0.0188	3502	1	AJ012609 Drosophila melanogast
gb_in1:AB011283	+ 395.50	190.25	0.0194	760	1	AB011283 Caenorhabditis elegans
gb_pr4:AK026441	+ 394.50	182.91	0.0496	2813	1	AK026441 Homo sapiens cDNA: FT
gb_pr7:HSINSP4BP	+ 381.00	177.37	0.1010	2837	1	X89399 Homo sapiens mRNA for I

gb\_ro:MMU20238 + 381.00 175.41 0.1298 4107 ! U20238 Mus musculus GTPase-  
gb\_in1:AB011285 + 378.50 182.84 0.0500 833 ! AB011285 Caenorhabditis eleg  
gb\_ro:AF086714 + 371.00 172.80 0.1814 3117 ! AF086714 Mus musculus rasc  
gb\_ro:AF055883 + 360.50 182.79 0.0504 211 ! AF055883 Rattus norvegicus p

seq\_name: gb\_ro:AF050183

seq\_documentation\_block:  
LOCUS AF050183 4265 bp mRNA ROD 15-SEP-2000  
DEFINITION Rattus norvegicus GTPase activating protein SynGAP-c mRNA, complete  
cds.

ACCESSION AF050183  
VERSION AF050183.2 GI:10140854

KEYWORDS

SOURCE Norway rat.

ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 4265)

AUTHORS Kim, J.H., Liao, D., Lau, L.F. and Haganir, R.L.

TITLE SynGAP: a synaptic RasGAP that associates with the PSD-95/SAP90

JOURNAL protein family

MEDLINE Neuron 20 (4), 683-691 (1998)

REFERENCE 2 (bases 1 to 4265)

AUTHORS Kim, J.H. and Haganir, R.L.

TITLE Direct Submission

JOURNAL

COMMENT Submitted (20-FEB-1998) Neuroscience, Johns Hopkins Medical

Institute, 725 N. Wolfe Street, PCTB 900, Baltimore, MD 21205, USA

On Sep 15, 2000 this sequence version replaced gi:3044054.

FEATURES

Location/Qualifiers

1..4265

/organism="Rattus norvegicus"

/db\_xref="taxon:10116"

/tissue\_type="hippocampus"

/dev\_stage="adult"

277..3792

/note="neuronal synaptic RasGAP; associates with the

PSD-95/SAP90 protein family; alternatively spliced

variants, SynGAP-a and SynGAP-b"

/codon\_start=1

/product="GTPase activating protein SynGAP-c"

/protein\_id="AAC40082.2"

/db\_xref="GI:10140855"

/translation="MQSEKESHESLSPSSAAEALELNLEDEDSIIKPVHSSILGQE

FCFEVTTSSGTCFACSAERDKWLNQRAVKPNKDNRRVNDVNLKLIIEARELP

PKKRYICELCLDDMLYARTTSKPRASGDTVFGEHFEFNNLPVAVRALRLHLYRDSK

KKKDKAGYVGLTVPVATLAGRHFEQWYPTLTPTGGSGGSGMGSGGGSGGGSGG

KGKGGCPAVRLKARYQTMILPMELKFEAEYVTHYMLCAVLEPALNVKKEEVA

ALVHILQSTGKAKDFLSDMAMSEVDREMERHLEFRNTLCAKIEEYMRLLIGQKYLK

DAIGFIRALYSEENECVDPKCTASSIAEHQALNMCCELALCKVNVSHCVFPREL

KEVFASWRLRCAREGEDIAIRLISASLFLRFLCPAIMSPSLFGLMQEYPDEOTSRTL

YIDLRELSTHALLWEVLPQLSKEALKLGLPLRLSLSDISTALRNPNIQRPSRQSE

RARSQPMVLRGPSAEMQGYMMRDLSNSTDLOSFMARGLNSSMDMARLPSPTKEKPPPP

PPGGKDLFYVSRPLARSPAYCTSSSDITEPEQKMLSVNKSVMMLDLQDGPGRRL

NSSSVNLAAVGLDLSQASLTAALGLRPAPAGRLSQSGSSSITAAGMRLLSQMGVTT

DGVPAQQLRIPLSFQNPFLHMAADGPPAGHGGSSGHGPPSSHHHHHHHHHRRGGE

PGDTFAPFHGYSKSEDLSTGVKPPAAISILHSYSDSEFGSPGDTFTRRLQSLQDNLO

HMLSPQITIGPQPAPSGPGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG

SGNLLQSPESYGPAPRQQLSKESGSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG

PTMPASERTVAVSNMPLSADIESAHIEREYKLEKYSKSMDESLDRVKEVEEIH

SILKERHMSNRKLEEYERRLLSQEEQTSKILMQYQARLESEKRLRQQQVEKDSQIKS

ITIGRLMVEEELRDHPMAEPLPEPKRRLDLDAQVERQLPPLGPTNPRVTLPAPWNG

APPAPPPPRQLQITENGFEFRNTADH"

BASE COUNT 936 a 1358 c 1203 g 768 t

ORIGIN

alignment\_scores:

Quality: 6950.00

Ratio: 5.245

Length: 1328

Gaps: 3

Percent Similarity: 99.774 Percent Identity: 99.699

alignment\_block:

US-09-294-298-6 x AF050183 ..

Align seg 1/1 to: AF050183 from: 1 to: 4265

1 GlyGluThrGluLeuProGlnAlaProHisPheProPheAlaProGlnG1 17  
 97 GGAGAGACTGAGCTGCCCAAGCACCCCAATTCCCATTTGCTCCCCAGCA 146  
 17 nGlyPheLeuSerArgArgLeuLysSerSerIleLysArgThrLysSerG 34  
 147 AGGCTTCTGAGCGGAGGCTAAAAAGCTCCATCAAAAGCTACAAAGTCAC 196  
 34 InProLysLeuAspArgThrSerSerPheArgGlnIleLeuProArgPhe 50  
 197 AACCCAAACTTGACCGGACCGAGCTTTTCGACAGATCCTGCCTGCTTC 246  
 51 ArgSerAlaAspHisAspArgAlaArgLeuMetGlnSerPheLysGluSe 67  
 247 CGAAGTGCTGACCATGACCGGCGCGCTGATGCAGAGCTTCAAGGAGTC 296  
 67 rHisSerHisGluSerLeuLeuSerProSerSerAlaAlaGluAlaLeuG 84  
 297 TCACTCCCATGAGTCCCTGCTGAGTCCAGCAGTCTGCTGAGGCCCTGG 346  
 84 luLeuAsnLeuAspGluAspSerIleIleLysProValHisSerSerIle 100  
 347 AGCTCAACCTGGATGAAGACTCCATATCAAGCCAGCTACACAGCTCCATC 396  
 101 LeuGlyGlnGluPheCysPheGluValThrThrSerSerGlyThrLysCy 117  
 397 CTGGCCAGGAGTCTGCTTTGAGTAACAACATCGTCTGGGACAAATG 446  
 117 sPheAlaCysArgSerAlaAlaGluArgAspLysTrpIleGluAsnLeuG 134  
 447 TTTTGCTGCTGCTGCTGCGCGGAAAGGACAAATGGATGAGATCTAC 496  
 134 InArgAlaValLysProAsnLysAspAsnSerArgArgValAspAsnVal 150  
 497 AGAGGGCTGTGAACCCCAACAAGGACAAACAGCCGCGGCTAGATAACGTG 546  
 151 LeuLysLeuTrpIleIleGluAlaArgGluLeuProProLysLysArgTy 167  
 547 CTGAACCTATGATCATAGAAGCTCGAGAGCTGCCCCCAAGAGCGGATA 596  
 167 rTyrCysGluLeuCysLeuAspAspMetLeuTyrAlaArgThrThrSerL 184  
 597 TTACTGCGAGTTATGCTGAGCAGATGCTCTATGCAGGACCACTTCCA 646  
 184 ysProArgSerAlaSerGlyAspThrValPheTrpGlyGluHisPheGlu 200  
 647 AGCCCGCTCAGCCTCAGGAGACACTGTCTTTTGGGCGGAGCACTTCGAG 696  
 201 PheAsnAsnLeuProAlaValArgAlaLeuArgLeuHisLeuTyrArgAs 217  
 697 TTTAAACAACCTGCCTGCTGTCCGGGCGCTGCGGCTGCATCTGTACCGTGA 746  
 217 pSerAspLysLysArgLysLysAspLysAlaGlyTyrValGlyLeuValT 234  
 747 CTCGGACAAAAGCGGAAGAGGACAAAGGACAGGCTACGTTGGCCTGGTGA 796  
 234 hrValProValAlaThrLeuAlaGlyArgHisPheThrGluGlnTrpTyr 250  
 797 CTGTTCCAGTGCCACCCCTGGCTGGGCGGCCACTTCACAGAGCAGTGGTAC 846  
 251 ProValThrLeuProThrGlySerGlySerGlyGlyMetGlySerG1 267  
 847 CCCGTGACCCCTGCCAACAGGAAGTGGGGCTCTGGGGGTATGGGCTCGGG 896  
 GlyGlyGlyGlySerGlyGlySerGlyGlySerGlyGlyLysGlyGlyC 284

897 GGGAGGAGGGGGTTCAGGGGGGGGGCTCAGGGGGGCAAGGAAAGGAGGCT 946  
 284 ysProAlaValArgLeuLysAlaArgTyrGlnThrMetSerIleLeuPro 300  
 947 GTCCTGCTGTGCGGCTGAAGGCCCGTTACCAGACATGAGTATCCTGCC 996  
 301 MetGluLeuTyrLysGluPheAlaGluTyrValThrAsnHisTyrArgMe 317  
 997 ATGGAGCTATATAAGGAGTTTGCAGAATATGTGACCAACCACTACCGCAT 1046  
 317 tLeuCysAlaValLeuGluProAlaLeuAsnValLysGlyLysGluGluV 334  
 1047 GCTGTGTGCGGCTGCTGGAGCCCGCTCAATGTCAAGGGCAAGGAGGAG 1096  
 334 alAlaSerAlaLeuValHisIleLeuGlnSerThrGlyLysAlaLysAsp 350  
 1097 TCGCTAGTGCACCTGGTTTCACATCTCTGCAAAAGCACAGGCAAGGCCAAGGAC 1146  
 351 PheLeuSerAspMetAlaMetSerGluValAspArgPheMetGluArgG1 367  
 1147 TFCCTTTCAGACATGGCCATGTGAGAGGTAGACCGGTTCATGGAGCGGGA 1196  
 367 uHisLeuIlePheArgGluAsnThrLeuAlaThrLysAlaIleGluGluT 384  
 1197 ACACCTCATATTCGCGAGAACACAGCTCGCCACTAAAGCCATAGAAGAGT 1246  
 384 yrMetArgLeuIleGlyGlnLysTyrLeuLysAspAlaIleGlyGluPhe 400  
 1247 ATATGAGACTGATGGCCAGAAATACCTCAAGGATGCCATTTGGGAGTTC 1296  
 401 ileArgAlaLeuTyrGluSerGluGluAsnCysGluValAspProIleLy 417  
 1297 ATCCGGGCTCTGTATGAATCTGAGGAGAACTGTGAAGTAGACCCCATCAA 1346  
 417 sCysThrAlaSerSerLeuAlaGluHisGlnAlaAsnLeuArgMetCysC 434  
 1347 GTGCACAGCTCCAGTCTGGCAGAGCACCCAGGCCCACTGCGGATGTGCT 1396  
 434 ysGluLeuAlaLeuCysLysValValAsnSerHisCysValPheProArg 450  
 1397 GTGAGTTGGCCCTGTGCAAGGTGGTCAACTCCCAATTGCGTGTTCGCCGAG 1446  
 451 GluLeuLysGluValPheAlaSerTrpArgLeuArgCysAlaGluArgG1 467  
 1447 GAGCTGAAGGAGGTGTTTGCAATCATGCGCGCTGCGCTGTGCAGAGCGGG 1496  
 467 yArgGluAspIleAlaAspArgLeuIleSerAlaSerLeuPheLeuArgP 484  
 1497 CCGGGAGGACATTTGCTGACAGGCTGATCAGCGCTCGCTCTTCTTCTGCGCT 1546  
 484 heLeuCysProAlaIleMetSerProSerLeuPheGlyLeuMetGlnGlu 500  
 1547 TCCTCTGCCCGCCCATCATGTGCGCCAGTCTGTTTGGACTGATGCAGGAG 1596  
 501 TyrProAspGluGlnThrSerArgThrLeuThrLeuIleAlaLysValI1 517  
 1597 TACCCAGATGAGCAGACCTCACGAACCCCTCACCTCATCGCCAAGGTTAT 1646  
 517 eGlnAsnLeuAlaAsnPheSerLysPheThrSerLysGluAspPheLeuG 534  
 1647 CCAGAACCTGCCCAACTTTTCCAAGTTTACCTCAAAAGGAGGACTTCTCTGG 1696  
 534 lyPheMetAsnGluPheLeuGluLeuGluTrpGlySerMetGlnGlnPhe 550  
 1697 GCTTCATGAACGAGTTCCTGGAGCTGGAATGGGGTTCATGCAGCAATTC 1746  
 551 LeuTyrGluIleSerAsnLeuAspThrLeuThrAsnSerSerPheG1 567  
 1747 TTGTATGAGATATCCAACCTGGACACACTGACCAACAGCAGCAGTTCGTA 1796  
 567 uGlyTyrIleAspLeuGlyArgGluLeuSerThrLeuHisAlaLeuLeuT 584  
 1797 GGGCTACATAGACTTGGGCGCGGAGCTCTCCACACTTCACGCCCTGCTCT 1846

584 rpGluValLeuProGlnLeuSerLysGluAlaLeuLeuLysLeuGlyPro 600  
 |||||  
 1847 GGGAGGTGCTGCCCCAGCTCAGCAAGGAAGCCCTCCTGAAGCTGGGCCG 1896  
  
 601 LeuProArgLeuLeuSerAspIleSerThrAlaLeuArgAsnProAsnIl 617  
 |||||  
 1897 CTGCCCCGGCTCCTCAGCGACATCAGCACAGCCCTGAGGAACCTTAACAT 1946  
  
 617 eGlnArgGlnProSerArgGlnSerGluArgAlaArgSerGlnProMetV 634  
 |||||  
 1947 CCAAAGGCAGCCGAGCCCGCAGAGCGGCGCTCGGTCTCAGCCCATGG 1996  
  
 634 alIeuArgGlyProSerAlaGluMetGlnGlyTyrMetMetArgAspLeu 650  
 |||||  
 1997 TGCTGCGGGCGCTCAGCCGAAATGCAGGGCTACATGATGCGGGACCTC 2046  
  
 651 AsnSerSerIleAspLeuGlnSerPheMetAlaArgGlyLeuAsnSerSe 667  
 |||||  
 2047 AACAGCTCCATCGACCTTCAGTCTTCATGGCTCGAGGCCTCAACAGCTC 2096  
  
 667 rMetAspMetAlaArgLeuProSerProThrLysGluLysProProProp 684  
 |||||  
 2097 TATGACATGGCTCGCCTCCCTCCTCCCAACCAAGGAGAAACCCCGCGC 2146  
  
 684 roProProGlyGlyLysAspLeuPheTyrValSerArgProProLeu 700  
 |||||  
 2147 CCCCCTCCCGGTGGGGTAAAGACCTGTTCTATGTGACCGGCCACCACTG 2196  
  
 701 AlaArgSerSerProAlaTyrCysThrSerSerSerAspIleThrGluPr 717  
 |||||  
 2197 GCCCGGTCTCCCCAGCATACTGCACGAGCAGCTCGACATCACAGAGCC 2246  
  
 717 oGluGlnLysMetLeuSerValAsnLysSerValSerMetLeuAspLeuG 734  
 |||||  
 2247 GGAGCAGAAGATGCTGAGTGTCAACAAGAGTGTGTCCATGCTGGACCTGC 2296  
  
 734 lnGlyAspGlyProGlyGlyArgLeuAsnSerSerSerValSerAsnLeu 750  
 |||||  
 2297 AGGGCAGCGGGCCTGGGGCGCCCTTAACAGCAGTAGTGTTCCAACCTG 2346  
  
 751 AlaAlaValGlyAspLeuLeuHisSerSerGlnAlaSerLeuThrAlaAl 767  
 |||||  
 2347 GCAGCTGTTGGGACCTGTTGCACCTCAAGCCAGGCTTCAC TGACAGCAGC 2396  
  
 767 aLeuGlyLeuArgProAlaProAlaGlyArgLeuSerGlnGlySerGlys 784  
 |||||  
 2397 CTTGGGGTTGGCGCCTGCACCTGCCGGCGCCCTCTCCCAAGGGAGTGGCT 2446  
  
 784 erSerIleThrAlaAlaGlyMetArgLeuSerGlnMetGlyValThrThr 800  
 |||||  
 2447 CTTCCATCACAGCAGCCGGCATGGCCCTCAGCCAGATGGGTGTCACTACG 2496  
  
 801 AspGlyValProAlaGlnGlnLeuArgIleProLeuSerPheGlnAsnPr 817  
 |||||  
 2497 GATGGTGTCCCCGCCAGCAACTGCGCATCCCTCTTTCTCTCCAGAACCC 2546  
  
 817 oLeuPheHisMetAlaAlaAspGlyProGlyProProAlaGlyHisGlyG 834  
 |||||  
 2547 TCTCTCCATATGGTGGCGATGGACCAGGCCCCCAGCGGCCCATGGAG 2596  
  
 834 lySerSerGlyHisGlyProProSerSerHisHisHisHisHisHis 850  
 |||||  
 2597 GGAGCAGTGGCCATGGTCCACCTTCCTCCCATCACCACCACCACCACCAT 2646  
  
 851 HisHisHisArgGlyGlyGluProProGlyAspThrPheAlaProPheHi 867  
 |||||  
 2647 CACCATACCGAGGGGAGAACCCCGAGGGGACACTTTTGGCCCCGTTCCA 2696  
  
 867 sGlyTyrSerLysSerGluAspLeuSerThrGlyValProLysProProA 884  
 |||||  
 2697 TGGCTATAGCAAGAGCGAGGACCTCTCTACAGGGGTCCCTAAGCCCCCTG 2746

884 laAlaSerIleLeuHisSerHisSerTyrSerAspGluPheGlyProSer 900  
 |||||  
 2747 CGGCCCTCCATCCTTCACAGCCACAGCTACAGTGATGAGTTGGACCCCTCT 2796  
  
 901 GlyThrAspPheThrArgArgGlnLeuSerLeuGlnAspAsnLeuGlnHi 917  
 |||||  
 2797 GGTACTGATTTTACCCGTGGCAGCTCTCACTTCAGGACAACCTACAGCA 2846  
  
 917 sMetLeuSerProProGlnIleThrIleGlyProGlnArgProAlaProS 934  
 |||||  
 2847 CATGCTCTCCCGCCCCAGATCACCATCGGTCCCCAGAGGCCAGTCCCT 2896  
  
 934 erGlyProGlyGlySerGlyGlySerGlySerGlyGlyGlyGlyGly 950  
 |||||  
 2897 CAGGCCCAGGAGGGGCAGTGGTGGGGCAGTGGTGGGGCGGTGGGGGC 2946  
  
 951 GlnProProLeuGlnArgGlyLysSerGlnGlnLeuThrValSerAl 967  
 |||||  
 2947 CAGCCACCTCCCTTG CAGAGGGGGCAAATCTCAGCAGTTGACAGTGAGTGC 2996  
  
 967 aAlaGlnLysProArgProSerSerGlyAsnLeuLeuGlnSerProGluP 984  
 |||||  
 2997 TGCCCAAGAAACCCCGCCGTCCTCCAGCGGGAACCTATTGCAGTCCCCGGAAC 3046  
  
 984 roSerTyrGlyProAlaArgProArgGlnGlnSerLeuSerLysGluGly 1000  
 |||||  
 3047 CAAGTTATGGTCTGCCCGTCCACGGCAACAGAGCCTCAGCAAGAGAGGGC 3096  
  
 1001 SerIleGlyGlySerGlyGlySerGlyGlyGlyGlyGlyGlyGlyLeuLy 1017  
 |||||  
 3097 AGCATTGGGGCAGCGGGGCAGCGGTGGCGGAGGGGTGGGGGCTCAA 3146  
  
 1017 sProSerIleThrLysGlnHisSerGlnThrProSerThrLeuAsnProT 1034  
 |||||  
 3147 GCCCTCCATCACCAAGCAGCATTC CCAGACTCCATCCACGCTGAACCCCA 3196  
  
 1034 hrMetProAlaSerGluArgThrValAlaTrpValSerAsnMetProHis 1050  
 |||||  
 3197 CGATGCCGGCCTCGGAGCGGACTGTAGCCTGGGTGTCCAATATGCTCTAC 3246  
  
 1051 LeuSerAlaAspIleGluSerAlaHisIleGluArgGluGluTyrLysLe 1067  
 |||||  
 3247 CTGTCCGCTGACATCGAGAGTGCACACATTGAGCGGGAAGAGTACAAGCT 3296  
  
 1067 uLysGluTyrSerLysSerMetAspGluSerArgLeuAspArgValLysG 1084  
 |||||  
 3297 GAAGGAGTACTCGAAGTCCATGGACGAGAGCCGACTGGACAGGGTGAAGG 3346  
  
 1084 luTyrGluGluGluIleHisSerLeuLysGluArgLeuHisMetSerAsn 1100  
 |||||  
 3347 AGTACGAGGAGGAGATCCACTCACTGAAGAAAGGCTACACATGTCCAAC 3396  
  
 1101 ArgLysLeuGluGluTyrGluArgArgLeuLeuSerGlnGluGlnTh 1117  
 |||||  
 3397 CGGAAGCTGGAAGAGTACGAGCGGAGGCTGCTGTCCCAGGAAGAGCAGAC 3446  
  
 1117 rSerLysIleLeuMetGlnTyrGlnAlaArgLeuGluGlnSerGluLysA 1134  
 |||||  
 3447 CAGCAAGATCCTGATGCAGTACC AAGCCCGCCTGGAGCAGAGCGAGAAGC 3496  
  
 1134 rgLeuArgGlnGlnValGluLysAspSerGlnIleLysSerIlelle 1150  
 |||||  
 3497 GCTTGAGGCAGCAGCAGGTGGAGAGGACTCCCAGATCAAGAGCATCATT 3546  
  
 1151 GlyArgLeuMetLeuValGluGluLeuArgArgAspHisProAlaMe 1167  
 |||||  
 3547 GGCAGGCTGATGCTGGTGGAGGAGGAGTGGCGCGGACCACCCCGCCAT 3596  
  
 1167 tAlaGluProLeuProGluProLysLysArgLeuLeuAspAla.Gln... 1182  
 |||||  
 3597 GGCTGAGCCGCTGCCTGAACCCCAAGAGAGGCTGCTCGACGCTCAGGTGG 3646  
  
 1183 ArgGlySerPheProProTrpValGlnGlnThrArgVal...ArgTrpPr 1198



```
|||||
3647 AGAGGCAGCTCCCCCTTGGGTCCAAACAAACCCGCGTGTGACGCTGGCC 3696
|||||
1198 oHisLeuGlyThrAlaTrpProProGlnProHisProHisProGlyC 1215
|||||
3697 CCACCTTGAACGGCTGGCCCCCAGCCCAACCCCAACCCCGGCT 3746
|||||
1215 yArgSerGlnArgThrAlaSerSerGlyThrProGlnThrThrSerPro 1231
|||||
3747 GCAGATCACAGAACGGCGAGTTCGGAACACCCGACAGACCAAGCCCA 3796
|||||
1232 ProSerIleThrAspLeuLeuProCysAlaProTyrProGlyProProSe 1248
|||||
3797 CCCAGCATCACAGACCTCCTTCCCTGTGCACCCCTACCCCGGCCACCCAG 3846
|||||
1248 rValThrAspLeuLeuProSerAlaProAspProGlyThrSerProThrT 1265
|||||
3847 CGTCACAGACCTCCTTCCAGTGACCCGACCCCTGGAACATCACCAACCA 3896
|||||
1265 hrArgThrGlyArgHisGlnGlyThrAlaGlyLeuSerProLeuThrPro 1281
|||||
3897 CCAGGACTGGACGTCAACCAAGGACAGCGGATTGTCTCCCTTAACGCCT 3946
|||||
1282 ProTrpGlyThrHisLeuSerThrProLeuLeuHisSerArgArgGluSe 1298
|||||
3947 CCTTGGGACACCATCTGTCAACCCCACTGCTCCATTCCAGAGGGAGAG 3996
|||||
1298 rGlyThrLeuSerCysProLeuThrProGlyHisHisLeuProHisThra 1315
|||||
3997 TGGGACCTCAGCTGCCCTCTCACCCCAAGGACACCCACCTACCCCAACAG 4046
|||||
1315 spProPheThrLeuGlyCysTyrProHisPro 1325
|||||
4047 ACCCCTTCACTCTGGGGTGTATCCCCATCCT 4078
```

seq\_name: gb\_ro:AF058790

```
seq_documentation_block:
LOCUS AF058790 4140 bp mRNA ROD 09-OCT-1998
DEFINITION Rattus norvegicus SynGAP-b mRNA, complete cds.
ACCESSION AF058790
VERSION AF058790.1 GI:3722228
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 4140)
AUTHORS Kim,J.H., Liao,D., Lau,L.F. and Huganir,R.L.
TITLE SynGAP: a synaptic RasGAP that associates with the PSD-95/SAP90
protein family
JOURNAL Neuron 20 (4), 683-691 (1998)
MEDLINE 98240917
REFERENCE 2 (bases 1 to 4140)
AUTHORS Kim,J.H. and Huganir,R.L.
TITLE Direct Submission
JOURNAL Submitted (09-APR-1998) Neuroscience, Johns Hopkins Medical
Institute, 725 N. Wolfe Street, PCTB 900, Baltimore, MD 21205, USA
REFERENCE 3 (bases 1 to 4140)
AUTHORS Kim,J.H. and Huganir,R.L.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1998) Neuroscience, Johns Hopkins Medical
Institute, 725 N. Wolfe Street, PCTB 900, Baltimore, MD 21205, USA
REMARK Sequence update by submitter
COMMENT On Oct 9, 1998 this sequence version replaced gi:3065890.
FEATURES
source
1..4140
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_type="hippocampus"
1..3750
/note="neuronal RasGAP; contains PH domain, C2 domain, GAP
```

```
domain in the N-terminal region, and a T/SXV motif at the
C-terminus"
/codon_start=1
/product="SynGAP-b"
/protein_id="AAC63511.1"
/db_xref="GI:3722229"
/translation="MGLRPPTPTPSGGSGSLPPPSHRQPLRRRCSSCCFPGEYHLG
RSRRKSVPGGKQYSMEAAPAFRPSQGFSLRRRLKSKRTKSPKLDRTSSFRQILP
RERSADHDRARLMQSFKESHESLSPSSAAEALELNDEDSIIPKVHSSILQEEFC
FEVTTSSGTKCFACAAERDKWIENLQRAVKPNKDNRRVDNLKWIIEARELPPK
KRYYCELCDDMLYARTTSKPRSASGDTVFGEHFEFNNLPVAVRALRLHLYRSDKKR
KKDKAGYVGLTVPVATLAGRHFTQWYPVTLPTSGSGSGMGSGGGSGGGSGGKG
KGGCPAVRLKARYQTMSILPMELYKEFAEYVTHYRMLCAVLEPALNVKKEEVASAL
VHILOSTGKAKDFLSDMAMSEVDRFMEREHLIFRENTLATKAIIEYMRLIGOKYLKDA
IGEFIRALYSEENCEVDPIKCTASSLAEOANLRCCELALCKVNVNSHCVPRELKE
VFASWRLRCAERGEDIDRLISASLFLRFLCPAITSPLFLGLMOEYPDEQTSRTLTL
IAKVIONLANFSKTFKEDFLGFMEFLFLEWGSQQOFLYEISNLDTLTNSSSFEGYI
DLGRELSTLHALLWEVLPLSKAALLKGLPLRLSDISTALRNPNIQRPQRSERA
RSQPMVLRGPSAEMQGYMMRDNLSSIDLQSEMARGLNSSMDMARLPSPPTKEPPPPP
GGKDLFVYVRPPLARSSPAYCTSSSDITEPEOKMLSVNKSVMLDLOGDGPGRRLNS
SSVSNLAAVGDLLHSSQASLTAALGRPAPAGRI.SQGGSSITAAAGMRLSQMGVTTDG
SVAQQLRIPLSFQNLFFHMAADGPGPPAGHGSGSGHGPSSHHHHHHHHHRRGEPGP
DTFAFFHGYSKSEDLTVGPKPPAAAILHSYSDEFPSGDTDFTRRLSLQDNLQHM
LSPQITIGPQRPAPSGPGGGSGGGSGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
NLLQSPESYGPAPRPQSLKEGSIIGSGGGSGGGGGGGGGGGGGGGGGGGGGGGGG
MPASERTVAVSNMPLSADIESAHIEREYKKEYSKSMDESRLDRVKEYEIEIHSI
KERLHMSNRKLEERYERLLSQEEQTSKILMOYQARLEQSEKRLRQQQVEKDSQIKSII
GRMLVEEELRRDHPAMAEPLPEPKRRLDAAQRGSFPPVWQQTRV"
```

BASE COUNT 902 a 1328 c 1182 g 728 t  
ORIGIN

alignment\_scores:

Quality: 6912.50 Length: 1319  
Ratio: 5.257 Gaps: 1  
Percent Similarity: 99.697 Percent Identity: 99.621

alignment\_block:

US-09-294-298-6 x AF058790 ..

Align seg 1/1 to: AF058790 from: 1 to: 4140

```
8 AlaProHisPheProPheAlaProGlnGlnGlyPheLeuSerArgArgLe 24
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 GCCCGCGCTGCCCTTCCGGCCTCGAAGGCTTCCTGAGCCGAGGCT 233

24 uLysSerSerIleLysArgThrLysSerGlnProLysLeuAspArgThrs 41
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
234 AAAAGCTCCATCAAAACGTACAAAGTCACAAACCACTTGACCGGACCA 283

41 erSerPheArgGlnIleLeuProArgPheArgSerAlaAspHisAspArg 57
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
284 GCAGCTTCGACAGATCCTGCCTCGCTTCGGAAGTGTGACCATGACCGG 333

58 AlaArgLeuMetGlnSerPheLysGluSerHisSerHisGluSerLeuLe 74
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
334 GCCCGCTGATGCAGAGCTTCAAGGAGTCTCACTCCCATGAGTCCCTGCT 383

74 uSerProSerAlaAlaGluAlaLeuGluLeuAsnLeuAspGluAsps 91
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
384 GAGTCCCAGCAGTGTGCTGAGGCCCTGGAGCTCAACCTGGATGAAGACT 433

91 erIleIleLysProValHisSerSerIleLeuGlyGlnGluPheCysphe 107
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
434 CCATTATCAAGCCAGTACACAGCTCCATCCTGGGCCAGGAGTTCTGCTT 483

108 GluValThrThrSerSerGlyThrLysCysPheAlaCysArgSerAlaAl 124
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
484 GAGGTAACAACATCGTCTGGGACAAAATGTTTTCCTGCTCGGTCTGCAGC 533

124 aGluArgAspLysTrpIleGluAsnLeuGlnArgAlaValLysProAsnL 141
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
534 CGAAAGGGACAAATGGATTGAGAAATCTACAGAGGGCTGTGAAACCCCAACA 583
```



141 ysAspAsnSerArgArgValAspAsnValLeuLysLeuTrpIleLeuGlu 157  
158 AlaArgGluLeuProProLysLysArgTyrCysGluLeuLysLeuAs 174  
634 GCTCGAGAGTGCCTCCCAAGAGCGATATTAAGTGGAGTTATGGCTGGA 683  
174 pAspMetLeuTyrAlaArgThrThrSerLysProArgSerAlaSerGlyA 191  
684 CGACATGCTATATGCACGGACCACTTCCAAGCCCGCTCAGCCCTCAGGAG 733  
191 spThrValPheTrpGlyGluHisPheGluPheAsnAsnLeuProAlaVal 207  
734 ACACCTGCTTTGGGGCGGAGCACTTCGAGTTAAACAACCTGCCCTGCTGTC 783  
208 ArgAlaLeuArgLeuHisLeuTyrArgAspSerAspLysLysArgLysLy 224  
784 CGGGCGCTGGGCTGCATCTGTACCGTGTACCCCTGACCTGCCAACAGAA 833  
224 sAspLysAlaGlyTyrValGlyLeuValThrValProValAlaThrLeuA 241  
834 GGACAAGGCGAGCTACGTTGGCTGGTACTGTTCAGTGGCCACCTGG 883  
241 laGlyArgHisPheThrGluGlnTrpTyrProValThrLeuProThrGly 257  
884 CTGGCGCCACTTCACAGAGCACTGGTACCCCTGACCTGCCAACAGGA 933  
258 SerGlyGlySerGlyGlyMetGlySerGlyGlyGlyGlySerGlyGly 274  
934 AGTGGGGGCTTGGGGGTATGGGCTCGGGGGGAGGAGGGGGTACGGGGG 983  
274 yGlySerGlyGlyLysGlyLysGlyGlyCysProAlaValArgLeuLysA 291  
984 CGGCTCAGGGGGCAAGGAAAGGAGGCTGTCTCTGCTGTGGGCTGAAGG 1033  
291 laArgTyrGlnThrMetSerIleLeuProMetGluLeuTyrLysGluPhe 307  
1034 CCCGTTACCAAGACAATGAGTATCTCTGCCCATGGAGCTATATAAGAGTTT 1083  
308 AlaGluTyrValThrAsnHisTyrArgMetLeuLysAlaValLeuGluPr 324  
1084 GCAGAATATGTGACCAACCACTACCGCATGCTGTGTGCTGCTGGAGCC 1133  
324 oAlaLeuAsnValLysGlyLysGlyGluValAlaSerAlaLeuValHisI 341  
1134 CGCCCTCAATGTCAAGGGCAAGGAGGAGGTCGCTAGTGCATGCTGCTACA 1183  
341 leLeuGlnSerThrGlyLysAlaLysAspPheLeuSerAspMetAlaMet 357  
1184 TCCTGCAAGACAGGCAAGGCAAGGACTTCTCTTCAGACATGSCCATG 1233  
358 SerGluValAspArgPheMetGluArgGluHisLeuIlePheArgGluAs 374  
1234 TCAGAGGTAGACCGGTTTCATGGAGCGGGAACACCTCATATTCGCGAGAA 1283  
374 nThrLeuAlaThrLysAlaIleGluGluTyrMetArgLeuIleGlyGlnL 391  
1284 CACGCTCGCCACTAAAGCCATAGAGAGTATATGAGACTGATTGGCCAGA 1333  
391 ysTyrLeuLysAspAlaIleGlyGluPheIleArgAlaLeuTyrGluSer 407  
1334 AATACCTCAAGGATGCCATTGGGGAGTTTCATCCGGGCTCTGTATGAATCT 1383  
408 GluGluAsnCysGluValAspProIleLysCysThrAlaSerSerLeuAl 424  
1384 GAGGAGAACTGTGAAGTAGACCCCATCAAGTGCACAGCGTCCAGTCTGGC 1433  
424 aGluHisGlnAlaAsnLeuArgMetCysCysGluLeuAlaLeuCysLysV 441  
1434 AGAGCACCAAGGCCAACCTGCGGATGTGCTGTGAGTTGGCCCTGTGCAAGG 1483

441 aiValAsnSerHisCysValPheProArgGluLeuLysGluValPheAla 457  
1484 TGGTCAACTCCCATTCGCTGTTCCGAGGGAGCTGAAGGAGGTGTTGCA 1533  
458 SerTrpArgLeuArgCysAlaGluArgGlyArgGluAspIleAlaAspAr 474  
1534 TCATGGCGGCTGCGCTGTGCAGAGCGGGCGGAGGACATTCGTCGACAG 1583  
474 gLeuIleSerAlaSerLeuPheLeuArgPheLeuCysProAlaIleMets 491  
1584 GCTGATCAGCGCTCGCTCTTCTCGCTTCTCTGCGGCGCCATCATGT 1633  
491 erProSerLeuPheGlyLeuMetGlnGluTyrProAspGluGlnThrSer 507  
1634 CGCCAGTCTGTGGACTGATGCAGGAGTACCCAGATGAGCAGACCTCA 1683  
508 ArgThrLeuThrLeuIleAlaLysValIleGlnAsnLeuAlaAsnPhese 524  
1684 CGAACCTTCACCTCATCGCAAGGTATCCAGAACCTGGCCCACTTTTC 1733  
524 rLysPheThrSerLysGluAspPheLeuGlyPheMetAsnGluPheLeuG 541  
1734 CAAGTTTACCTCAAAGGAGGACTTCTTGGGCTTCAATGAACGAGTTCTGG 1783  
541 luLeuGluTrpGlySerMetGlnGlnPheLeuTyrGluIleSerAsnLeu 557  
1784 AGCTGGAATGGGTTCTATGCAGCAATTTGTATGAGATATCCAACCTG 1833  
558 AspThrLeuThrAsnSerSerSerPheGluGlyTyrIleAspLeuGlyAr 574  
1834 GACACACTGACCAACAGCAGCAGTTTGGGGCTACATAGACTTGGGCGG 1883  
574 gGluLeuSerThrLeuHisAlaLeuLeuTrpGluValLeuProGlnLeuS 591  
1884 CGAGCTCTCCACACTTCACGCCCTGCTTCTGGAGGTGCTGCCCCAGCTCA 1933  
591 erLysGluAlaLeuLeuLysLeuGlyProLeuProArgLeuLeuSerAsp 607  
1934 GCAAGGAAGCCCTCTGAAGCTGGGCGGCTGCTGCCCGGCTCCTCAGCGAC 1983  
608 IleSerThrAlaLeuArgAsnProAsnIleGlnArgGlnProSerArgG1 624  
1984 ATCAGCACAGCCCTGAGGAACCCCTAACATCCAAAGGAGCGGAGCGGCA 2033  
624 nSerGluArgAlaArgSerGlnProMetValLeuArgGlyProSerAlaG 641  
2034 GAGCGAGCGGCTCGGCTCAGCCCATGCTGCTGCGGCGGCGCTCAGCGG 2083  
641 luMetGlnGlyTyrMetMetArgAspLeuAsnSerSerIleAspLeuGln 657  
2084 AAATGCAGGGCTACATGATGCGGGACCTCAACAGCTCCATCGACCTTCAG 2133  
658 SerPheMetAlaArgGlyLeuAsnSerSerMetAspMetAlaArgLeuPr 674  
2134 TCCTTCATGGCTCGAGGCTCAACAGCTCTATGGACATGGCTCGCCTCCC 2183  
674 oSerProThrLysGluLysProProProProProGlyGlyGlyLysA 691  
2184 CTCCCAACCAAGAGAGAAACCCCGGCGGCTCCCGGTTGGGGTAAAG 2233  
691 spLeuPheTyrValSerArgProProLeuAlaArgSerSerProAlaTyr 707  
2234 ACCTGTTCTATGTGAGCGGCGCCACCTGCGGCTCCCTCCCGCAGCATAC 2283  
708 CysThrSerSerSerAspIleThrGluProGluGlnLysMetLeuSerVa 724  
2284 TGCACGAGCAGCTCGGACATCACAGAGCGGAGCAGAGATGCTGAGTGT 2333  
724 lAsnLysSerValSerMetLeuAspLeuGlnGlyAspGlyProGlyGlyA 741  
2334 CAACAAGAGTGTGTCATGCTGGACCTGCAGGGCGGCGGCTGGGGGCC 2383  
741 rgLeuAsnSerSerSerValSerAsnLeuAlaAlaValGlyAspLeuLeu 757

|||||  
2384 GCCTTAACAGCAGTAGTGTTCCTCAACCTGGCAGCTGTGGGAGCCTGTTG 2433  
758 HisSerSerGlnAlaSerLeuThrAlaAlaLeuGlyLeuArgProAlaPr 774  
|||||  
2434 CACTCAAGCCAGGCTTCACTGACAGCAGCTTGGGGTGGGGCTGCACC 2483  
774 oAlaGlyArgLeuSerGlnGlySerGlySerSerIleThrAlaAlaGlyM 791  
|||||  
2484 TGCCGGCGCCTCTCCCAAGGGAGTGGCTTCCATCACAGCAGCCGCA 2533  
791 etArgLeuSerGlnMetGlyValThrThrAspGlyValProAlaGlnGln 807  
|||||  
2534 TGCGCTCAGCCAGATGGGTGTCACTACGATGGTGTCCCGCCAGCAA 2583  
808 LeuArgIleProLeuSerPheGlnAsnProLeuPheHisMetAlaAlaAs 824  
|||||  
2584 CTGCGCATCCCTCTTCTCTCCAGAACCTCTCTCCATATGGCTGCCGA 2633  
824 pGlyProGlyProProAlaGlyHisGlyGlySerSerGlyHisGlyProp 841  
|||||  
2634 TGGACAGGGCCCCCAGAGGCTTGGAGGAGCAGTGGCCATGGTCCAC 2683  
841 roSerSerHisHisHisHisHisHisHisHisHisHisHisHisHisHis 857  
|||||  
2684 CTTCTCCCATCACCAACCACCATCACCATCACCATCACCATCACCATCAC 2733  
858 ProGlyAspThrPheAlaProPheHisGlyTyrSerLysSerGluAs 874  
|||||  
2734 CCCCAGGGGACACTTTGGCCCGTTCATGGCTATAGCAGAGCGAGGA 2783  
874 pLeuSerThrGlyValProLysProProAlaAlaSerIleLeuHisSerH 891  
|||||  
2784 CCTCTCTACAGGGGTCCCTAAGCCCTTGGGCTTCCATCTCCATCTCACAG 2833  
891 isSerTyrSerAspGluPheGlyProSerGlyThrAspPheThrArgArg 907  
|||||  
2834 ACAGTACAGTGTAGTGTGGACCTCTGGTACTGATTTACCCGCTCGG 2883  
908 GlnLeuSerLeuGlnAspAsnLeuGlnHisMetLeuSerProGlnI 924  
|||||  
2884 CAGCTCTCACTTCAGGACAACTTACAGCACATGCTCTCCCCGCCACAGAT 2933  
924 eThrIleGlyProGlnArgProAlaProSerGlyProGlyGlySerG 941  
|||||  
2934 CACCATCGTCCCCAGAGCCAGCTCCCTCAGGGCCAGGAGGGGCGAGTG 2983  
941 lyGlyGlySerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 957  
|||||  
2984 GTGGGGCAGTGGTGGGGGGGTGGGGGGCCAGCCACCTCCCTTGCAGAGG 3033  
958 GlyLysSerGlnGlnLeuThrValSerAlaAlaGlnLysProArgProse 974  
|||||  
3034 GGCAATCTCAGCAGTTGACAGTGAGTGCTGCCAGAAACCCCGCCGTC 3083  
974 rSerGlyAsnLeuLeuGlnSerProGluProSerTyrGlyProAlaArgP 991  
|||||  
3084 CAGCGGAACCTATTGAGTCCCCGGAACCAAGTTATGGTCTCTCCCGTC 3133  
991 roArgGlnGlnSerLeuSerLysGlyGlySerIleGlyGlySerGlyGly 1007  
|||||  
3134 CACGCAACAGAGCCTCAGCAAGAGGGCAGCATTTGGGGCAGCGGGGGC 3183  
1008 SerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyH 1024  
|||||  
3184 AGCGGTGGGGGGGGGGGGGGCTCAGCCCTCCATCACCAAGCAGCA 3233  
1024 sSerGlnThrProSerThrLeuAsnProThrMetProAlaSerGluArgT 1041  
|||||  
3234 TTCCAGACTCCATCCAGCTGAACCCACGATGCCGGCTCGGAGCGGA 3283  
rValAlaAlaTrpValSerAsnMetProHisLeuSerAlaAspIleGluSer 1057  
|||||

3284 CTGTAGCCTGGGTGTCCAATATGCCTCACCTGTCCGCTGACATCGAGAGT 3333  
1058 AlaHisIleGluArgGluGluTyrLysLeuLysGluTyrSerLysSerMe 1074  
|||||  
3334 GCACACATTGAGCGGGAAGAGTACAAGCTGAAGGAGTACTCGAAGTCCAT 3383  
1074 tAspGluSerArgLeuAspArgValLysGluTyrGluGluIleHis 1091  
|||||  
3384 GGACGAGAGCCGACTGGACAGGGTGAAGAGTACGAGGAGAGATCCACT 3433  
1091 erLeuLysGluArgLeuHisMetSerAsnArgLysLeuGluTyrGlu 1107  
|||||  
3434 CACTGAAGGAAAGGCTACACATGTCCAAACCGGAAGCTGGAAGAGTACGAG 3483  
1108 ArgArgLeuLeuSerGlnGluGluGlnThrSerLysIleLeuMetGlnTy 1124  
|||||  
3484 CGGAGGCTGTGTCCAGGAAGAGAGCAGACCAAGATCCTGATGCAGTA 3533  
1124 rGlnAlaArgLeuGluGlnSerGluLysArgLeuArgGlnGlnValG 1141  
|||||  
3534 CCAAGCCCGCTGGAGCAGAGCGAGAGCGCTTGAGGCAGCAGAGGTGG 3583  
1141 luLysAspSerGlnIleLysSerIleGlyArgLeuMetLeuValGlu 1157  
|||||  
3584 AGAAGGACTCCCATCAAGAGCATCATTTGGCAGGCTGATGGTGGAG 3633  
1158 GluGluLeuArgArgAspHisProAlaMetAlaGluProLeuProGluPr 1174  
|||||  
3634 GAGGAGCTGCGCCGAGACCCCGCCATGGCTGAGCCGCTGCCTGAACC 3683  
1174 olLysLysArgLeuLeuAspAlaGlnArgGlySerPheProTrpValG 1191  
|||||  
3684 CAAGAAGAGGCTGCTCGACGCTCAGAGAGCAGCTTCCCCCTTGGGTCC 3733  
1191 lnGlnThrArgVal...ArgTrpProHisLeuGlyThrAlaTrpProPro 1206  
|||||  
3734 AACAAACCCGCTGTGACGCTGGCCCCACCTTGGAAACGGCTTGGCCCCC 3783  
1207 GlnProHisProHisProGlyCysArgSerGlnArgThrAlaSerSe 1223  
|||||  
3784 CAGCCCCACCCCGCCCGCCGCTGCAGATCACAGAGAACGGCGAGTTC 3833  
1223 rGlyThrProGlnThrThrSerProProSerIleThrAspLeuLeuProC 1240  
|||||  
3834 CGGAACACCGCAGACCACTAGCCACCCAGCAGCATCACAGACCTCCTTCCCT 3883  
1240 ysAlaProTyrProGlyProProSerValThrAspLeuLeuProSerAla 1256  
|||||  
3884 GTGACCCCTACCCCGCCCGCCAGCGCTCAGAGACCTCCTTCCCAGTGCA 3933  
1257 ProAspProGlyThrSerProThrThrArgThrGlyArgHisGlnGlyTh 1273  
|||||  
3934 CCGACCCCTGGAACATCACCAACCAACCAAGACTGGACGTACCAAGGAC 3983  
1273 rAlaGlyLeuSerProLeuThrProProTrpGlyThrHisLeuSerThrP 1290  
|||||  
3984 AGCGGATTTGCTCCTTAACGCTTCTTGGGGCACCCCATCTGTCAACCC 4033  
1290 roLeuLeuHisSerArgArgGluSerGlyThrLeuSerCysProLeuThr 1306  
|||||  
4034 CACTGCTCCATTCCAGGAGGAGAGTGGGACCCCTCAGCTGCCCTCTCAC 4083  
1307 ProGlyHisHisLeuProHisThrAspProPheThrLeuGlyCysTyrPr 1323  
|||||  
4084 CCAGGACACCACTTACCCACACAGACCCCTTCACTCTGGGGTGTATCC 4133  
1323 oHisPro 1325  
|||||  
4134 CCATCCT 4140  
seq\_name: gb\_ro:AF058789  
seq\_documentation\_block:





1216 GCAGAAATATGTGACCAACCACTACCGCATGCTGTGTGGCTGTGGAGCC 1265  
324 oAlaLeuAsnValLysGlyLysGluValAlaSerAlaLeuValHisI 341  
1266 CGCCCTCAATGTCAAGGCAAGGAGGAGTGGCTAGTGCATGGTTTACA 1315  
341 leLeuGlnSerThrGlyLysAlaLysAspPheLeuSerAspMetAlaMet 357  
1316 TCCTGCAAGACACAGGCAAGGCCAAGGACTTCCTTTTCAGACATGGCCATG 1365  
358 SerGluValAspArgPheMetGluArgGluHisLeuIlePheArgGluAs 374  
1366 TCAGAGGTAGACCGGTTTCATGGAGCGGGAACACCTCATATTCGCGAGAA 1415  
374 nThrLeuAlaThrLysAlaIleGluGluTyrMetArgLeuIleGlyGlnL 391  
1416 CACGCTCGCCACTAAAGCCATAGAGAGTATATGAGACTGATGGCCAGA 1465  
391 yTyrLeuLysAspAlaIleGlyGluPheIleArgAlaLeuTyrGluSer 407  
1466 AATACCTCAAGGATGCCATTTGGGAGTTTCATCCGGGCTCTGTATGAATCT 1515  
408 GluGluAsnCysGluValAspProIleLysCysThrAlaSerSerLeuAl 424  
1516 GAGGAGAACTGTGAAGTAGACCCCATCMAGTGCACAGCGTCCAGTCTGGC 1565  
424 aGluHisGlnAlaAsnLeuArgMetCysCysGluLeuAlaLeuCysLysV 441  
1566 AGAGCACCAAGGCCAACCTGCGGATGTGCTGTGAGTTGGCCCTGTGCAAGG 1615  
441 alValAsnSerHisCysValPheProArgGluLeuLysGluValPheAla 457  
1616 TGGTCAACTCCCATTTGGCTGTTCCCGAGGAGCTGAAGGAGGTGTTTGA 1665  
458 SerTrpArgLeuArgCysAlaGluArgGlyArgGluAspIleAlaAspAr 474  
1666 TCATGGCGCTGCGCTGTGCAGAGCGGGCGGAGGACATGCTGCACAG 1715  
474 gLeuIleSerAlaSerLeuPheLeuArgPheLeuCysProAlaIleMets 491  
1716 GCTGATCAGCGCTCGCTTCTCTCGGCTTCTCTGCGGCGCCATCATGT 1765  
491 erProSerLeuPheGlyLeuMetGlnGluTyrProAspGluGlnThrSer 507  
1766 CGCCAGTGTGTTGGACTGATGCAGGAGTACCCAGATGAGCAGACCTCA 1815  
508 ArgThrLeuThrLeuIleAlaLysValIleGlnAsnLeuAlaAsnPhese 524  
1816 CGAACCTTACCCTCATCCCAAGGTTATCCAGAACCTGGCCAACTTTTC 1865  
524 rLysPheThrSerLysGluAspPheLeuGlyPheMetAsnGluPheLeuG 541  
1866 CAAGTTTACCTCAAAGGAGGACTTCTGCGGCTTCATGAACGAGTTTCTGG 1915  
541 luLeuGluTrpGlySerMetGlnGlnPheLeuTyrGluIleSerAsnLeu 557  
1916 AGCTGAGTGGGGTTCTATGCAGCAATCTGTGATGAGATATCCAACCTG 1965  
558 AspThrLeuThrAsnSerSerPheGlyGlyTyrIleAspLeuGlyAr 574  
1966 GACACACTGACCAACACAGCAGGTTTGGAGGCTACATAGACTTGGCCCG 2015  
574 gGluLeuSerThrLeuHisAlaLeuLeuTrpGluValLeuProGlnLeuS 591  
2016 CGAGCTCTCCACACTTCAGCCCTGCTCTGGGAGGTGCTGCCCCAGCTCA 2065  
591 erLysGluAlaLeuLeuLysLeuGlyProLeuProArgLeuLeuSerAsp 607  
2066 GCAAGGAAGCCCTCCTGAAGCTGGGCGCGCTGCCCGGCTCCTCAGCGAC 2115  
608 ileSerThrAlaLeuArgAsnProAsnIleGlnArgGlnProSerArgG1 624  
2116 ATCAGCACAGCCCTGAGGAACCCCTAACATCCAAAGGACGCGGAGCCGCCA 2165

624 nSerGluArgAlaAargSerGlnProMetValLeuArgGlyProSerAlaG 641  
2166 GAGCGAGCGCGCTCGGCTCAGCCCATGCTGCTGCGCGGCGCTCAGCGG 2215  
641 luMetGlnGlyTyrMetArgAspLeuAsnSerSerIleAspLeuGln 657  
2216 AGATGCAAGGGCTACATGATCGGGACCTCAACAGTCCATCGACCTTCAG 2265  
658 SerPheMetAlaArgGlyLeuAsnSerSerMetAspMetAlaArgLeuPr 674  
2266 TCCTTCATGGCTCGAGGCTCAACAGCTCTATGGACATGGCTCGCCTCCC 2315  
674 oSerProThrLysGluLysProProProProProProGlyGlyLysA 691  
2316 CTCCCCAACCAAGGAGAAACCCCGCGCCCTCCCGTGGGGTAAAG 2365  
691 spLeuPheTyrValSerArgProProLeuAlaArgSerSerProAlaTyr 707  
2366 ACCTGTTCTATGTAGCGGGCCACCACTGGCCCGTCCCTCCACGACATAC 2415  
708 CysThrSerSerSerAspIleThrGluProGluGlnLysMetLeuSerVa 724  
2416 TGCACGACGAGCTCGGACATCACAGCGCGGAGCAGAGATGCTGAGTGT 2465  
724 lAsnLysSerValSerMetLeuAspLeuGlnGlyAspGlyProGlyGlyA 741  
2466 CAACAAGAGTGTGTCCATGTGGACCTGCAGGGCGGCGGCTGGGGGCC 2515  
741 rgLeuAsnSerSerSerValSerAsnLeuAlaAlaValIcilyAspLeuLeu 757  
2516 GCCTTAACAGCAGTAGTGTTCCTCAACCTGGCAGCTGTGGGGACCTGTG 2565  
758 HisSerSerGlnAlaSerLeuThrAlaAlaLeuGlyLeuArgProAlaPr 774  
2566 CACTCAAGCCAGGCTTCACTGACAGCAGCTTGGGTTGCGGCTGCACC 2615  
774 oAlaGlyArgLeuSerGlnGlySerGlySerSerIleThrAlaAlaGlyM 791  
2616 TGCGGGCGGCTCTCTCCCAAGGAGTGGCTCTTCCATCACAGCAGCGGCA 2665  
791 etArgLeuSerGlnMetGlyValThrThrAspGlyValProAlaGlnGln 807  
2666 TGCGCTCAGCCAGATGGGTGTCACTACGATGGTGTCCCGCCAGCAA 2715  
808 LeuArgIleProLeuSerPheGlnAsnProLeuPheHisMetAlaAlaAs 824  
2716 CTGCGCATCCCTCTTCTTCCAGAACCTCTTCCATATGCTGCTGCGA 2765  
824 pGlyProGlyProProAlaGlyHisGlyGlySerSerGlyHisGlyProp 841  
2766 TGGACCGGCGGCCCCAGCGGCTTGGAGGAGCAGTGGCCATGCTGCCAC 2815  
841 roSerSerHisHisHisHisHisHisHisHisHisHisHisHisHisHis 857  
2816 CTCTCTCCATCACACACACACACACACACACACATCACCGAGGGGAGAA 2865  
858 ProProGlyAspThrPheAlaProPheHisGlyTyrSerLysSerGluAs 874  
2866 CCCCCAGGGGACACTTTTGGCCCGTTCCTATGGCTATAGCAAGAGCGAGGA 2915  
874 pLeuSerThrGlyValProLysProProAlaAlaSerIleLeuHisSerH 891  
2916 CCTCTACAGGGGTCTTCAAGCCCCCTGGGCTTCCATCTTCCATCTTCCAGCC 2965  
891 isSerTyrSerAspGluPheGlyProSerGlyThrAspPheThrArgArg 907  
2966 ACAGCTACAGTGTATGTTGGACCCCTCTGGTACTGATTTTACCCGCTCG 3015  
908 GlnLeuSerLeuGlnAspAsnLeuGlnHisMetLeuSerProGlnIle 924  
3016 CAGCTCTCACTTCAGGACAAACCTACAGCAGCATGCTCTCCCGGCGCCAGAT 3065

924 eThrIleGlyProGlnArgProAlaProSerGlyProGlyGlyGlySerG 941  
|||||  
3066 CACCATCGTCCCGAGAGGCCAGCTCCCTCAGGGCCAGGAGGGGGCAGTG 3115  
  
941 lyGlyGlySerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 957  
|||||  
3116 GTGGGGCAGTGGTGGGGCGGTGGGGCCAGCCACCTCCCTTGCAGAGG 3165  
  
958 GlyLysSerGlnGlnLeuThrValSerAlaAlaGlnLysProArgProSe 974  
|||||  
3166 GGCAATCTCAGCAGTTGACAGTGAGTGCTGCCAGAAACCCCGGCCGTC 3215  
  
974 rSerGlyAsnLeuLeuGlnSerProGluProSerTyrGlyProAlaArgP 991  
|||||  
3216 CAGCGGAACCTATTGCAATGCCCGGAACCAAGTTATGCTGCTGCCGTC 3265  
  
991 roArgGlnGlnSerLeuSerLysGluGlySerIleGlyGlySerGlyGly 1007  
|||||  
3266 CACGGCAACAGAGCCCTCAGCAAGAGGGCAGCATTTGGGGGCAGCGGGGC 3315  
  
1008 SerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1024  
|||||  
3316 AGCGGTGGCGGAGGGGTGGGGGCTCAAGCCCTCCATCACCAAGCAGCA 3365  
  
1024 sSerGlnThrProSerThrLeuAsnProThrMetProAlaSerGluArgT 1041  
|||||  
3366 TTCCAGACTCCATCCACGCTGAACCCACGATGGCGGCTCGGAGCGGA 3415  
  
1041 hrValAlaTrpValSerAsnMetProHisLeuSerAlaAspIleGluSer 1057  
|||||  
3416 CTGTAGCTGGGTGTCATATGCTCCACCTGCTGCTGACATCGAGAGT 3465  
  
1058 AlaHisIleGluArgGluGluTyrLysLeuLysGluTyrSerLysSerMe 1074  
|||||  
3466 GCACACATTGAGCGGGAAGAGTACAAAGTGAAGAGTACTCGAAGTCCAT 3515  
  
1074 tAspGluSerArgLeuAspArgValLysGluTyrGluGluIleHisS 1091  
|||||  
3516 GGACGAGAGCCGACTGGACAGGTGAAGGAGTACGAGGAGGAGATCCACT 3565  
  
1091 erLeuLysGluArgLeuHisMetSerAsnArgLysLeuGluGluTyrGlu 1107  
|||||  
3566 CACTGAAGAAAGGCTACACATGTCCAACCGGAAGTGAAGAGTACGAG 3615  
  
1108 ArgArgLeuLeuSerGlnGluGlnThrSerLysIleLeuMetGlnTy 1124  
|||||  
3616 CGGAGGCTGTCTCCAGGAAGAGCAGACCCAGCAAGATCCTGTATGCACTA 3665  
  
1124 rGlnAlaArgLeuGluGlnSerGluLysArgLeuArgGlnGlnValG 1141  
|||||  
3666 CCAAGCCGCTGGAGCAGAGCGAGAAGCGCTTGAGGCAGCAGCAGGTGG 3715  
  
1141 luLysAspSerGlnIleLysSerIleIleGlyArgLeuMetLeuValGlu 1157  
|||||  
3716 AGAAGGACTCCAGATCAAGAGCATCATTTGGCAGGCTGATGCTGGTGGAG 3765  
  
1158 GluGluLeuArgArgAspHisProAlaMetAlaGluProLeuProGluPr 1174  
|||||  
3766 GAGGAGCTGCGCGGACCAACCCCGCATGGCTGAGCCGCTGCTGTAACC 3815  
  
1174 oLysLysArgLeuLeuAspAlaGln..... 1182  
|||||  
3816 CAAGAGAGGCTGCTCGACGCTCAGCTCCTCATCAGGTAATTCTCCTGCT 3865  
  
1182 ..... 1182  
  
3866 TCCGCTTTGACCACGGCGGAGGACACTGGGGGAGGTGACTCCGGACCAC 3915  
  
1183 .....ArgGlySerPheProTrpValGlnGlnThrArgVal...Ar 1196  
|||||  
3916 TGCAGGAGAGGAGCAGCTTCCCCCTTGGGTCCCAACAACCCCGGTGTGACG 3965  
  
1196 gTrpProHisLeuGlyThrAlaTrpProProGlnProHisProProHisP 1213

|||||  
3966 CTGGCCCCACCTTGAACGGCTGGCCCCCCCCCAGCCCCCCCCCACC 4015  
  
1213 roGlyCysArgSerGlnArgThrAlaSerSerGlyThrProGlnThrThr 1229  
|||||  
4016 CCGGCTGCAGATCACAGAGAACGGCGAGTCCGGAACACCGCAGACCACT 4065  
  
1230 SerProProSerIleThrAspLeuLeuProCysAlaProTyrProGlyPr 1246  
|||||  
4066 AGCCACCCAGCATCACAGACCTCTTCCCTGTGCACCTTACCCCGGCC 4115  
  
1246 oProSerValThrAspLeuLeuProSerAlaProAspProGlyThrSerP 1263  
|||||  
4116 ACCCAGCTCACAGACCTCTTCCAGTGCACCCGACCCCTGGAACATCAC 4165  
  
1263 roThrThrArgThrGlyArgHisGlnGlyThrAlaGlyLeuSerProLeu 1279  
|||||  
4166 CAACCAACAGGAGTGGACGTCAACCAAGGAGACGCGGATTGTCTCCCTTA 4215  
  
1280 ThrProProTrpGlyThrHisLeuSerThrProLeuLeuHisSerArgAr 1296  
|||||  
4216 AGCCTCTCTTGGGCGACGCT.CTGTCAACCCCACTGTCTCCATCCAGGAG 4264  
  
1296 gGluSerGlyThrLeuSerCysProLeuThrProGlyHisHisLeuProH 1313  
|||||  
4265 GGAGAGTGGGACCTCAGCTGCCCTCTCACCCAGGACACCACTACCTACCC 4314  
  
1313 isThrAspProPheThrLeuGlyCysTyrProHisPro 1325  
|||||  
4315 ACACAGACCCCTCAGCTGGGTGCTATCCCCATCCT 4352

seq\_name: gb\_ro:AF048976

seq\_documentation\_block:

LOCUS AF048976 4063 bp mRNA

DEFINITION Rattus norvegicus synaptic ras GTPase-activating protein pl35

ACCESSION AF048976

VERSION AF048976.1 GI:2935447

KEYWORDS

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 4063)

AUTHORS Chen, H.-J. and Kennedy, M.B.

TITLE Identification and cloning of a novel 130 kd protein containing a

ras GTPase-activating domain from the rat forebrain postsynaptic

density

JOURNAL (in) SOC. NEUROSCI. ABSTR.: 1466; (1997)

REFERENCE 2 (bases 1 to 4063)

AUTHORS Chen, H.-J., Rojas-Soto, M. and Kennedy, M.B.

TITLE A synaptic Ras GTPase-activating protein (pl35 SynGAP) inhibited by

Cam kinase II

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 4063)

AUTHORS Chen, H.-J. and Kennedy, M.B.

TITLE Direct Submission

JOURNAL Submitted (17-FEB-1998) Division of Biology, California Institute

of Technology, 1200 E. California Blvd., MC 216-76, Pasadena, CA

91125, USA

FEATURES

Location/Qualifiers

1..4063

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

122..4003

/function="prominent substrate for endogenous CamKII; can

activate intrinsic ras GTPase activity"

/note="synaptic ras-GAP; N-terminal encodes putative PH

domain, C2 domain, and ras-GAP domain; C-terminal encodes

proline-rich region, stretch of 10 histidine residues, and

Beckman Institute  
6

27-MAR-1998

```
t/sxv motif; enriched in the forebrain postsynaptic
density fraction"
/codon_start=1
/product="synaptic ras GTPase-activating protein p135
SynGAP"
/protein_id="AAC08071.1"
/db_xref="GI:2935448"
/translation="MSYAPFRDVRGPPMHRTQYVHSPYDRPGWNPFRFCIIISGNQLLML
DEDEIHPLIIRDRSESSRNKLLRTVSVPEGRPHGEHYHLGRSRKSPYGGKQYS
MEAPAAPFRPSQGFSLRRKSSIKRTKSQPKLDRSSFRQILPRFSADHRRARLMQ
SFKESHSHESLSPSSAAEALELNLDSDSIKPVHSSILGQEFCEVTTSSGTKCFAC
RSAERDKWIENLQRAVKPNKDNRRVDNVLKWIIEARELPKKRYCYCELCDDMLY
ARTTSKPRSAGDITVFWGEHFEFNNLPVAVRALRLHLRYRSDKKRKKDKAGYVGLVTP
VATLAGRHFTQWYPTVLTPTGSGGSGMGSGGGGSGGSGGSGGKGGCPAVRLKARYQ
TMSILPMELYKEFAEYVTNHYRMLCAVLEPALNVKGEVASALVHILQSTGKAKDFL
SDMAMSEVDREMEREHLLIFRENTLATAKIEEYMRLLIGOKYLKDAIGEFIRALYESEN
CEVDPIKCTASSLAHQANLRMCCELALCKVNVSHCVFPRELKEVFASWRLRCAERGR
EDIAADRLISASLFLRFLCPAIMSPSLFGLMQEYPDEQTSRTLTLLIAKVIQNLANSKF
TSKEDFLGFNNEFLEWGSMMQFLYEISNLDTLTNSSSFEYIDLGRELSTLHALLW
EVLPLQSKAELKGLPLRLLSDISTALRNPNIQRSQSERARSRQPMVLRGPSAEM
QGYMMRDNLSSIDLQSFMARGLNSMDMARLPSPTEKPPPPPPGGGKDLFYVSRPPL
ARSSPAYCTSSSDITEPEQKMLSVNKSVMMLDQGDGPGRLNSSSVSNLAAGVDLLH
SSQASLTALGLRPAGRLSQSGSSITAAGMRLSQMGVTTDGVPAQQLRPLPSFQN
PLFMAADGGPPAGHGGSSGHPSSHHHHHHHHHRRGEGEPGDTFAPPHGYSKSED
LSTGVPKPPAASILHSHYSDEFPSGDTFTRRQLSQDNLOHMLSPPTIGPQRPAP
PSGPGGSGGGGGGGGGQPPPLQKSKSOQLTVSAAQKPRPSSGNLLQSPSPSYGPAR
PQQSLSKESIGSGGGGGGGGLKPSITKQHSQTPSTLNPMPASERTVAVWSNM
PHLSADIESAHIEREYKLKEYSKSMDESRLDRKEYEEDIEHSLKERLHMSNRKLEEY
ERRLLSQEEQTSKILMQYQARLEQSEKRLRQQQVEKDSQIKSIIGRLMLVEEELRRDH
PMAEPLPEPKRLLDAQRGSPFPWVQTRV"
BASE COUNT      880 a  1273 c  1177 g  733 t
ORIGIN
```

```
alignment_scores:
  Quality: 6292.50      Length: 1209
  Ratio: 5.222          Gaps: 1
Percent Similarity: 99.669 Percent Identity: 99.586

alignment_block:
US-09-294-298-6 x AF048976 ..
Align seg 1/1 to: AF048976 from: 1 to: 4063
      8 AlaProHisPhePropheAlaProGlnGlnGlyPheLeuSerArgArgLe 24
      437 GCGCCCGCTGCGCCCTTCGCGCCCTCGCAAGGCTTCTGAGCGCGAGGCT 486
      24 uLysSerSerIleLysArgThrLysSerGlnProLysLeuAspArgThrS 41
      487 AAAAGCTCCATCAACAGTACAAAGTCACAAACCAACTTGACCGGACCA 536
      41 erSerPheArgGlnIleLeuProArgPheArgSerAlaAspHisAspArg 57
      537 GCAGCTTTCGACAGATCCTGCGCTCCGAAAGTGTGCTGACCATGACCGG 586
      58 AlaArgLeuMetGlnSerPheLysGluSerHisSerHisGluSerLeuLe 74
      587 GCGCGGCTGATGCAGAGCTTCAAGGAGTCTCACTCCCATGAGTCCCTGCT 636
      74 uSerProSerSerAlaAlaGluAlaLeuGluLeuAsnLeuAspGluAspS 91
      637 GAGTCCCAGCAGTGCTGTGCTGAGGCCCTTGAGCTCAACCTGGATGAAGACT 686
      91 erIleIleLysProValHisSerSerIleLeuGlyGlnGluPheCysPhe 107
      687 CCATTATCAAGCCAGTACACAGCTCCATCCTTGGGCGCAGGAGTTCTGCTT 736
      108 GluValThrThrSerSerGlyThrLysCysPheAlaCysArgSerAlaAl 124
      737 GAGGTAAACAACATCGTCTGGGACAAAATGTTTTCCTGCTGCTGCTGCAGC 786
      124 aGluArgAspLysTrpIleGluAsnLeuGlnArgAlaValLysProAsnL 141
```

```
||||| 787 CGAAAGGGACAAATGGATTGAGATCTACAGAGGGCTGTGAACCCCAACA 836
||||| 141 ysAspAsnSerArgArgValAspAsnValLeuLysLeuTrpIleIleGlu 157
||||| 837 AGGACAACAGCGCGGGTAGATAACGTGCTGAAACTATGATGATCATAGAA 886
||||| 158 AlaArgGluLeuProProLysLysArgTyrTyrCysGluLeuCysLeuAs 174
||||| 887 GCTCGAGAGCTGCCCCCAAGAAAGCGCATATTACTGCGAGTTATGCCTGGA 936
||||| 174 pAspMetLeuTyrAlaArgThrThrSerLysProArgSerAlaSerGlyA 191
||||| 937 CGACATGCTCTATGCACGACCACTTCCAAGCCCGCTCAGCCTCAGGAG 986
||||| 191 spThrValPheTrpGlyGluHisPheGluPheAsnAsnLeuProAlaVal 207
||||| 987 ACACGTGCTTTTGGGGCGGAGCACCTCGAGTTTAAACAACCTGCCTGCTGTC 1036
||||| 208 ArgAlaLeuArgLeuHisLeuTyrArgAspSerAspLysLysArgLysLy 224
||||| 1037 CGGCGCTGCGGCTGCATCTGTACCGTGACTCGGACAAAAAGCGGAAGA 1086
||||| 224 sAspLysAlaGlyTyrValGlyLeuValThrValProValAlaThrLeuA 241
||||| 1087 GGACAAGCGCAGGCTACGTTGGCCTGTGACTGTTCCAGTGCGCACCCCTGG 1136
||||| 241 laGlyArgHisPheThrGluGlnTrpTyrProValThrLeuProThrGly 257
||||| 1137 CTGGCGCGCACTTCACAGAGCAGTGTGTACCCCGTGACCCCTGCCAACAGGA 1186
||||| 258 serGlyGlySerGlyGlyMetGlySerGlyGlyGlyGlyGlyGlyGlyGly 274
||||| 1187 AGTGGGGGCTCTGGGGGTATGGGCTCGGGGGGAGGAGGGGGGTCAGGGGG 1236
||||| 274 yGlySerGlyGlyLysGlyLysGlyCysProAlaValArgLeuLysA 291
||||| 1237 CGGCTCAGGGGGCAAGGGAAAGGAGGCTGTCTGCTGTGCGGCTGAAGG 1286
||||| 291 laArgTyrGlnThrMetSerIleLeuProMetGluLeuTyrLysGluPhe 307
||||| 1287 CCCGTTTACCAGACAATGAGTATCCTGCCCATGGAGCTATATTAAGGAGTTT 1336
||||| 308 AlaGluTyrValThrAsnHisTyrArgMetLeuCysAlaValLeuGluPr 324
||||| 1337 GCAGAAATATGTACCAACCACTACCGCATGCTGTGTGCGGTGCTGGAGCC 1386
||||| 324 oAlaLeuAsnValLysGlyLysGluGluValAlaSerAlaLeuValHisI 341
||||| 1387 CGCCCTCAATGTCAAGGGCAAGGAGGAGGCTGCTAGTGCTGCTGCTTACA 1436
||||| 341 leLeuGlnSerThrGlyLysAlaLysAspPheLeuSerAspMetAlaMet 357
||||| 1437 TCCTGCAAAGCACAGGCAAGGCCAAGGACTTCTTTCAGACATGGCCATG 1486
||||| 358 SerGluValAspArgPheMetGluArgGluHisLeuIlePheArgGluAs 374
||||| 1487 TCAGAGGTAGACCGGTTTCATGGAGCGGGAACACCTCATATTCCCGGAGAA 1536
||||| 374 nThrLeuAlaThrLysAlaIleGluGluTyrMetArgLeuIleGlyGlnL 391
||||| 1537 CACGCTCGCCCACTAAAGCCCATAGAGACTATATGAGACTGATGGCCAGA 1586
||||| 391 ySTyrLeuLysAspAlaIleGlyGluPheIleArgAlaLeuTyrGluSer 407
||||| 1587 AATACCTCAAGGATGCCATTGGGGAGTTTCATCCGGGCTCTGTATGATCT 1636
||||| 408 GluGluAsnCysGluValAspProIleLysCysThrAlaSerSerLeuAl 424
||||| 1637 GAGGAGAACTGTGAAGTAGACCCCAAGTGCACAGCGTCCAGTCTGGC 1686
||||| 424 aGluHisGlnAlaAsnLeuArgMetCysCysGluLeuAlaLeuCysLysV 441
|||||
```



1687 AGAGCACAGGCCAACCTGCGGATGTGCTGTGAGTTGGCCCTGTGCAAGG 1736  
441 alValAsnSerHisCysValPheProArgGluLeuLysGluValPheAla 457  
1737 TGGTCAACTCCCATTCGGTGTCCCGAGGAGCTGAAGGAGGTGTGCA 1786  
458 SerTrpArgLeuArgCysAlaGluArgGlyArgGluAspIleAlaAspAr 474  
1787 TCATGGCGGCTGCGCTGTGCAGAGCGGGCGGAGGACATTCGTGACAG 1836  
474 gLeuIleSerAlaSerLeuPheLeuArgPheLeuCysProAlaIleMetS 491  
1837 GCTGATCAGCGCCTCGCTCTTCCTGCGCTTCCTGCGCCGCCCATCATGT 1886  
491 erProSerLeuPheGlyLeuMetGlnGluTyrProAspGluGlnThrSer 507  
1887 CGCCCATGCTGTTGGACTGATGCAGGAGTACCAGATGAGCAGACTCA 1936  
508 ArgThrLeuThrLeuIleAlaLysValIleGlnAsnLeuAlaAsnPheSe 524  
1937 CGAACCCCTACCCCTCATCGCCAAAGGTTATCCAGAACCTGGCCAACTTTC 1986  
524 rLysPheThrSerLysGluAspPheLeuGlyPheMetAsnGluPheLeuG 541  
1987 CAAGTTTACCTCAAAGGAGGACTTCCTGGGCTTCATGAACGAGTTTCTGG 2036  
541 luLeuGluTrpGlySerMetGlnGlnPheLeuTyrGluIleSerAsnLeu 557  
2037 AGCTGGAGTGGGTTCTATGCAGCAATTCCTGTATGAGATATCCAACTG 2086  
558 AspThrLeuThrAsnSerSerPheGluGlyTyrIleAspLeuGlyAr 574  
2087 GACACACTGACCAACACAGCAGCAGTTTGGAGGCTACATAGACTTGGGCGG 2136  
574 gGluLeuSerThrLeuHisAlaLeuLeuTrpGluValLeuProGlnLeuS 591  
2137 CGAGCTCTCCACACTTCACGCCCTGCTCTGGGAGGTGCTGCCCCAGCTCA 2186  
591 erLysGluAlaLeuLeuLysLeuGlyProLeuProArgLeuLeuSerAsp 607  
2187 GCAAGGAAGCCCTCCTGAAGCTGGGCCCGCTGCCCGGCTCCTCAGCGAC 2236  
608 IleSerThrAlaLeuArgAsnProAsnIleGlnArgGlnProSerArgGl 624  
2237 ATCAGCACAGCCCTGAGGAACCTTAACATCCAAAGCAGCGCAGCCGCCA 2286  
624 nSerGluArgAlaArgSerGlnProMetValLeuArgGlyProSerAlaG 641  
2287 GAGCGAGCGCGCTCGGTCTCAGCCCATGGTGTGCGCGGCGCTCAGCCG 2336  
641 luMetGlnGlyTyrMetMetArgAspLeuAsnSerSerIleAspLeuGln 657  
2337 AGATGACGGGCTACATGATGCGGGACCTCAACAGCTCTATGGACATGGCTCC 2386  
658 SerPheMetAlaArgGlyLeuAsnSerSerMetAspMetAlaArgLeuPr 674  
2387 TCCTTCATGGCTCGAGGCCCTCAACAGCTCTATGGACATGGCTCGCCCTCC 2436  
674 oSerProThrLysGluLysProProProProProProGlyGlyLysA 691  
2437 CTCCTCAACCAAGAGAAACCCCGCCCGCCCTCCCGGTTGGGGTAAAG 2486  
691 spLeuPheTyrValSerArgProProLeuAlaArgSerSerProAlaTyr 707  
2487 ACCTGTTCTATGTGAGCCGGCCACCATGTCGCCCGGTCCTCCCGCAGCATAC 2536  
708 CysThrSerSerAspIleThrGluProGluGlnLysMetLeuSerVa 724  
2537 TGCACGAGCAGCTCGGACATCACAGAGCCGGAGCAGAGAGATGCTGAGTGT 2586  
724 lAsnLysSerValSerMetLeuAspLeuGlnGlyAspGlyProGlyGlyA 741  
2587 CAACAAGAGTGTCTCCATGCTGGACCTGCAGGGCGGAGGGCCCTGGGGGCC 2636

741 rgLeuAsnSerSerSerValSerAsnLeuAlaAlaValGlyAspLeuLeu 757  
2637 GCCTTAACAGCAGTAGTGTTCCTCAACCTGGCAGCTGTGGGACCTGTTG 2686  
758 HisSerSerGlnAlaSerLeuThrAlaAlaLeuGlyLeuArgProAlaPr 774  
2687 CACTCAAGCCAGGCTTCACCTGACAGCAGCTTGGGTTGGGCTTGCAGC 2736  
774 oAlaGlyArgLeuSerGlnGlySerGlySerSerIleThrAlaAlaGlyM 791  
2737 TGCGGGGCGCTCTCCCAAGGAGTGGCTCTTCCATCACAGCAGCCGCA 2786  
791 etArgLeuSerGlnMetGlyValThrThrAspGlyValProAlaGlnGln 807  
2787 TGCGCTCAGCCAGATGGGTGTCACTACGATGGTGTCCCGCCAGCAA 2836  
808 LeuArgIleProLeuSerPheGlnAsnProLeuPheHisMetAlaAlaA 824  
2837 CTGCGCATCCCTCTTCTTCCAGAACCTCTCTTCCATATGGGTGCCGA 2886  
824 pGlyProGlyProProAlaGlyHisGlyGlySerSerGlyHisGlyProp 841  
2887 TGGACAGGGCCCCCAGCAGGCCATGGAGGAGCAGTGGCCATGGTCCAC 2936  
841 roSerSerHisHisHisHisHisHisHisHisHisHisHisHisHisHis 857  
2937 CTTCTCTCCCATCACCAACCAACCAACCAACCAACCAACCAACCAACCA 2986  
858 ProProGlyAspThrPheAlaProPheHisGlyTyrSerLysSerGluAS 874  
2987 CCCCAGGGGACACTTTTGGCCCGTTTCCATGGCTATAGCAAGAGCGAGGA 3036  
874 pLeuSerThrGlyValProLysProProAlaAlaSerIleLeuHisSerH 891  
3037 CCTCTCTACAGGGTCCCTAAGCCCCCTTGGGCTTCCATCTTCCATCTT 3086  
891 isSerTyrSerAspGluPheGlyProSerGlyThrAspPheThrArgArg 907  
3087 ACAGCTACAGTGTAGTGTGGACCTCTGTGTACTGATTTTACCCGTCGG 3136  
908 GlnLeuSerLeuGlnAspAsnLeuGlnHisMetLeuSerProGlnIl 924  
3137 CAGCTCTCACTTCAGGACAACCTACAGCACATGCTCTCCCGCCCCAGAT 3186  
924 eThrIleGlyProGlnArgProAlaProSerGlyProGlyGlySerG 941  
3187 CACCATCGGTCCCCAGAGGCCAGCTCCCTCAGGGCCAGGAGGGGCGAGTG 3236  
941 lyGlyGlySerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 957  
3237 GTGGGGCAGTGTGGGGGCGGTGGGGGCCAGCCACCTCCCTTGCAGAGG 3286  
958 GlyLysSerGlnGlnLeuThrValSerAlaAlaGlnLysProArgProse 974  
3287 GGCAAAATCTCAGCAGTTGACAGTGAGTGTGCCAGAAACCCCGCCGTC 3336  
974 rSerGlyAsnLeuLeuGlnSerProGluProSerTyrGlyProAlaArgP 991  
3337 CAGCGGGAACCTATTGTCAGTCCCGGGAACCACTATGTTCTTCCCGTC 3386  
991 roArgGlnGlnSerLeuSerLysGluGlySerIleGlyGlySerGlyGly 1007  
3387 CACGGCAACAGAGCCTCAGCAAGAGGGCAGGATGGGGGCGAGCGGGGGC 3436  
1008 SerGlyGlyGlyGlyGlyGlyGlyLeuLysProSerIleThrLysGlnHi 1024  
3437 AGCGGTGGCGGAGGGGTGGGGGCTCAAGCCCTCCATCACCACAGCAGCA 3486  
1024 sSerGlnThrProSerThrLeuAsnProThrMetProAlaSerGluArgT 1041  
3487 TTCCACAGACTCCATCCACGCTGAACCCCAACGATGCGGGCTCGGAGCGGA 3536

```
1041 hrValAlaTrpValSerAsnMetProHisLeuSerAlaAspLeuSer 1057
|||||
3537 CTGTAGCCTGGGTGTCATATGCCTCAGCTGCTCCGCTGACATCGAGAGT 3586
|||||
1058 AlaHisIleGluArgGluGluTyrLysLeuLysGluTyrSerLysSerMe 1074
|||||
3587 GCACACATTGACGGGAAGAGTACAAGCTGAAGGAGTACTCGAAGTCCAT 3636
|||||
1074 tAspGluSerArgLeuAspArgValLysGluTyrGluGluGluIleHiss 1091
|||||
3637 GGACGAGAGCCGACTGGACAGGCTGAAGGAGTACGAGGAGGAGATCCACT 3686
|||||
1091 erLeuLysGluArgLeuHisMetSerAsnArgLysLeuGluGluTyrGlu 1107
|||||
3687 CACTGAAGGAAAGGCTACACATGTCCAAACCGGAAGCTGGAAGAGTACGAG 3736
|||||
1108 ArgArgLeuLeuSerGlnGluGluGlnThrSerLysIleLeuMetGlnTy 1124
|||||
3737 CGAGGCTGCTGTCCAGGAAGAGCAGACGAGCAAGATCCTGATGCAGTA 3786
|||||
1124 rGlnAlaArgLeuGluGlnSerGluLysArgLeuArgGlnGlnValG 1141
|||||
3787 CCAAGCCCGCTGGAGCAGAGCGAGAGCGCTTGAGGAGCAGCAGGTGG 3836
|||||
1141 luLysAspSerGlnIleLysSerIleIleGlyArgLeuMetLeuValGlu 1157
|||||
3837 AGAAGGACTCCCAGATCAAGAGCATCATTGGCAGGCTGATGCTGGTGAG 3886
|||||
1158 GluGluLeuArgArgAspHisProAlaMetAlaGluProLeuProGluPr 1174
|||||
3887 GAGGAGCTGCGCCGGGACCAACCCGCCATGGCTGAGCCGCTGCTGAACC 3936
|||||
1174 oLysLysArgLeuLeuAspAlaGlnArgGlySerPheProProTrpValG 1191
|||||
3937 CAAGAAGAGGCTGCTCGACGCTCAGAGAGGAGCTTCCCCCTTGGGTCC 3986
|||||
1191 lnGlnThrArgVal...ArgTrpProHisLeuGlyThrAlaTrpProPro 1206
|||||
3987 AACAAACCCGCTGTGACGCTGGCCCCACCTTGGAACGGCTGGCCCCC 4036
|||||
1207 GlnProHisProHisProGlyCys 1215
|||||
4037 CAGCCCCACCCCCACCCCGGTGC 4063
|||||

seq_name: gb_ro:AB016962

seq_documentation_block:
LOCUS AB016962 4801 bp mRNA ROD 01-SEP-2000
DEFINITION Rattus norvegicus mRNA for synGAP-d, complete cds.
ACCESSION AB016962
VERSION AB016962.1 GI:4417206
KEYWORDS synGAP-bl.
SOURCE Rattus norvegicus (sub_species:Sprague Dawley) cDNA to mRNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 4801)
AUTHORS Suzuki,T.
TITLE SynGAP-d.
JOURNAL Published Only in DataBase (1999) In press
REFERENCE 2 (bases 1 to 4801)
AUTHORS Suzuki,T.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tatsuo
Suzuki, Shinshu University School of Medicine, Department of
Neuroplasticity; 3-1-1 Asahi, Matsumoto, Nagano 390-8621, Japan
(E-mail:suzukit@sch.md.shinshu-u.ac.jp, Tel:+81-263-37-2683,
Fax:+81-263-37-2725)
COMMENT On Mar 16, 1999 this sequence version replaced gi:4239945.
FEATURES
Sequence updated (09-Feb-1999).
Location/Qualifiers
source 1. .4801
```

```
/organism="Rattus norvegicus"
/sub_species="Sprague Dawley"
/db_xref="taxon:10116"
1124..4624
/gene="synGAP-d"
1124..4624
/gene="synGAP-d"
/note="synaptic ras GTPase activating protein"
/codon_start=1
/product="synGAP-d"
/protein_id="BAA74972.1"
/db_xref="GI:4417207"
/translation="MEYFPAAPFRPSQGLSRRLKSSIKRTKSQPKLDRSTSSFRQILP
RFRSADHDRARLMSQFKSHSHESLLSPSSAAEALELNDEDSIKPVHSSILGQEF
FEVTSSTGKCFACRSAAERDKWIENLQRAVKPNKDNRRVDNLKLIIEARELPPK
KRYCELCLDDMLYARTSKPRASGDTVFGEHFEFNNLPAVRALRLHLRYRSDKKR
KDKAGYVGLVTPVATLAGRHFTQWYPTVLTPTSGSGGSGGSGGSGGSGGSGG
KGGCPAVRLKARYQTMILPMELYKEFAEYVTHYRMLCAVLEPALNVKKEEVASAL
VHILQSTGKAKDFLSDMAMSEVDRFMRHLIFRENTLATKAI EYMRLIGQKYLKDA
IGEFIRALYSEENCEVDPIKCTASSLAHQANLRMCCELALCKVNVNSHCVFPRELKE
VFASWRLRCAERGEDIALRLISASLFLRFLCPAIMSPSLFLGLMQEYDPDEQTSRTLT
IAKVIQNLANFSKFTSKEDFLGFNFELEWGSQQOFLYEISNLDTLTNSSSEFYI
DLGRELSTLHALLWEVLPQLSKEALLKGLPLRLSDISTALRNPNIQRPQRSQSER
RSQPMVLRGPSAEMQGYMRDLNSSIDLQSPMARGNLSSMDMARLPSPTEKPPPPPP
GGKDLFYVSRPPLARSPAYCTSSDITEPEQKMLSVNKSVMISDLQGDGPGGRLNS
SSVSNLAAVGDLLHSSQASLTAAALRLPAPAGRLSQSGSSITAAAGMRLSQMGVTTDG
VPAQLRIPLSFQNPFLHMAADGPPGPHGSGSGSGGSGGSGGSGGSGGSGGSGG
DTFAPFHGYSKSEDLSTGVKPPAASILHSHSYSDGFGPSGDTDFTRQLSLQDNLOHM
LSPQITIGQRPAPSGPGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG
NLQSPSPSYGPARPROQSLKEGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG
MPASERTVAVSNMPLHSADIESAHIEREYKLEYSKSMDESRLDREYEEIHSKE
RLHMSNRKLEEYERLLSQEEQTSKILMOYQARLEQSEKRLRQOQVEKDSQIKSIIG
PSLQADAGGGGAAPGPRHG"
BASE COUNT 1043 a 1477 c 1314 g 967 t
ORIGIN

alignment_scores:
Quality: 6280.50 Length: 1225
Ratio: 5.178 Gaps: 5
Percent Similarity: 99.020 Percent Identity: 98.857

alignment_block:
US-09-294-298-6 x AB016962 ..
Align seg 1/1 to: AB016962 from: 1 to: 4801

9 ProHisPheProPheAlaProGlnGlnGlyPheLeuSerArgArgLeuLy 25
|||||
1136 CCGCTGCGCCCTTCGCGCCCTCGCAAGGCTCTCTGAGCCGAGGCTAAA 1185
|||||

25 sSerSerIleLysArgThrLysSerGlnProLysLeuAspArgThrSers 42
|||||
1186 AAGTCCCATCAACGTAACAAAGTACAAACCACTTGACCGACCAAGCA 1235
|||||

42 erPheArgGlnIleLeuProArgPheArgSerAlaAspHisArgAla 58
|||||
1236 GCTTCGACAGATCCTCGCTCGCTTCCGAAGTGTGACCATGACCGGGCC 1285
|||||

59 ArgLeuMetGlnSerPheLysGluSerHisSerHisGluSerLeuLeuSe 75
|||||
1286 CGGCTGATGCAGAGCTCAAGGAGTCTCACTCCCATGAGTCCCTGCTGAG 1335
|||||

75 rProSerSerAlaAlaGluAlaLeuGluLeuAsnLeuAspGluAspSerI 92
|||||
1336 TCCAGCAGTGTGTGAGGCCCTGGAGCTCAACCTGGATGAAGACTCCA 1385
|||||

92 leIleLysProValHisSerSerIleLeuGlyGlnGluPheCysPheGlu 108
|||||
1386 TTATCAAGCCAGTACACAGCTCCATCCTGGGCCAGGAGTTCTGCTTTGAG 1435
|||||

109 ValThrThrSerSerGlyThrLysCysPheAlaCysArgSerAlaAlaG1 125
|||||
```

425	uHisGlnAlaAsnLeuArgMetCysCysGluLeuAlaLeuCysLysValV	442
2386	GCACcAGGcCAAcCTGCGGATGTGCTGTGAGTTGGCCCTGTGCAAGGTGG	2435
442	alAsnSerHisCysValPheProArgGluLeuLysGluValPheAlaSer	458
2436	TCAAcTCcCATTCGCTGTTCCcGAGGcGAGcTGAAGGAGGTGTTTGATCA	2485
459	TrpArgLeuArgCysAlaGluArgGlyArgGluAspIleAlaAspArgLe	475
2486	TGGCGGCTCGcCTGTGCAGAGCGGGCCGGGAGcACATTGCTGACAGGCT	2535
475	uIleSerAlaSerLeuPheLeuArgPheLeuCysProAlaIleMetSerP	492
2536	GATCAGCGcCTCGcCTTCCTCGcGTTCCTCTGCGCGCCCATCATGTGCG	2585
492	roSerLeuPheGlyLeuMetGlnGluTyrProAspGluGlnThrSerArg	508
2586	CCAGTCTGTTGGACTGATGCAGGAGTAcCCAGATGAGCAGAcCTCACGA	2635
509	ThrLeuThrIleuIleAlaLysValIleGlnAsnLeuAlaAsnPheSerLy	525
2636	AcCCTCAcCCTCATCGcCAAGGTATCCAGAAcCTGGCCAACTTTTCcAA	2685
525	sPheThrSerLysGluAspPheLeuGlyPheMetAsnGluPheLeuGluL	542
2686	GTTTAcCTCAAGGAGGAcTTCTGGGcTTCATCAcGAGTTTCTTGAGc	2735
542	euGluTrpGlySerMetGlnGlnPheLeuTyrGluIleSerAsnLeuAsp	558
2736	TGGAGTGGGcTTCATGCAGcAAATTCTTGATGAGATATCCAACCTGGAc	2785
559	ThrLeuThrAsnSerSerSerPheGluGlyTyrIleAspLeuGlyArgG1	575
2786	AcACTGACCAAcAGCAGcAGTTTGAGGcCTAcATAGACTTGGGGCCCGA	2835
575	uLeuSerThrLeuHisAlaLeuLeuTrpGluValIleuProGlnLeuSerL	592
2836	GCTCTCCAcACTTCAGCGCCCTGCTCTGGGAGGTGCTGCCcCAGCTCAGcA	2885
592	ysGluAlaLeuLysLeuGlyProLeuProArgLeuLeuSerAspIle	608
2886	AGGAAGCCCTCTGAAGCTGGGGCCGCTGCCcCGcCTCTCAGCGAcATC	2935
609	SerThrAlaLeuArgAsnProAsnIleGlnArgGlnProSerArgGlnSe	625
2936	AGcACAGCCCTGAGGAACcCTTAACATCCAAAGGCAGCCGAGCCCGcAGAG	2985
625	rGluArgAlaArgSerGlnProMetValLeuArgGlyProSerAlaGluM	642
2986	CGAGCGCGCTCGGTCTcCAGCCCATGGTGCTGCGCGGGCCCGTCAGCCCGAG	3035
642	etGlnGlyTyrMetMetArgAspLeuAsnSerSerIleAspLeuGlnSer	658
3036	TGCAGGGCTAcATGATCGGGGAcCTCAcAGcTCTATGGAcATGGCTCGCCCTC	3085
659	PheMetAlaArgGlyLeuAsnSerSerMetAspMetAlaArgLeuProSe	675
3086	TTcATGGCTCGAGGGcCTCAAcAGcTCTATGGAcATGGCTCGCCCTCGCCCTC	3135
675	rProThrLysGluLysProProProProGlyGlyGlyLysAspL	692
3136	CCCAACCAAGAGAAACCCCGCGCCcCTCCCGGTGGGGGTAAAGAcC	3185
692	eupheTyrValSerArgProProLeuAlaArgSerSerProAlaTyrCys	708
3186	TGTTCTATGTAGCCGGCCAcCAcTGGCCCGGTCTCTCCcCAGCATAcTGC	3235
709	ThrSerSerSerAspIleThrGluProGluGlnLysMetLeuSerValAs	725
3236	AcGAGcAGCTCGGAcCATcCAGAGCCGGGAGcAGATGCTGAGTGTcAA	3285



725 nLysSerValSerMetLeuAspLeuGlnGlyAspGlyProGlyGlyArgL 742  
 3286 CAAGAGTGTGTCCATGCTGGACCTGCAGGGCGACGGGCGCTGGGGCGGCC 3335  
 742 euAsnSerSerSerValSerAsnLeuAlaValGlyAspLeuLeuHis 758  
 3336 TTAACAGCAGTAGTGTTCACACCTGGCAGCTGTTGGGGACCTGTTGCAC 3385  
 759 SerSerGlnAlaSerLeuThrAlaAlaLeuGlyLeuArgProAlaProAl 775  
 3386 TCAAGCCAGGCTTCACTGACAGCAGCCTTGGGGTTCGGGCTGCACCTGC 3435  
 775 aGlyArgLeuSerGlnGlySerGlySerSerIleThrAlaAlaGlyMeta 792  
 3436 CGGGCGCTCTCCCAAGGAGTGGCTCTTCCATCACAGCAGCGGGCATGC 3485  
 792 rgLeuSerGlnMetGlyValThrThrAspGlyValProAlaGlnGlnLeu 808  
 3486 GCCTCAGCCAGATGGGTGTCACTACGGATGGTGTCCCGCCAGCAACTG 3535  
 809 ArgIleProLeuSerPheGlnAsnProLeuPheHisMetAlaAlaAspG1 825  
 3536 CGCATCCCTCTTTCCTCCAGAACCTCTCTTCCATATGGCTGCCGATGG 3585  
 825 yProGlyProProAlaGlyHisGlyGlySerSerGlyHisGlyProProS 842  
 3586 ACCAGGGCCCCAGAGCCATGGAGGGAGCAGTGGCCATGGTCCACCTT 3635  
 842 erSerHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 858  
 3636 CCTCCATCACACCACCACCACCACCATCACCATCACCGAGGGGAGAACCC 3685  
 859 ProGlyAspThrPheAlaProPheHisGlyTyrSerLysSerLysLeu 875  
 3686 CCAGGGGACACTTTTGGCCCGTTCATGGCTATAGCAAGAGCGAGGACCT 3735  
 875 uSerThrGlyValProLysProProAlaAlaSerIleLeuHisSerHisS 892  
 3736 CTCTACAGGGTCCCTAAGCCCCCTGGGGCTCCATCCTTCCAGCCACA 3785  
 892 erTyrSerAspGluPheGlyProSerGlyThrAspPheThrArgArgGln 908  
 3786 GCTACAGTATGAGTTGGACCTCTGGTACTGATTTTACCCCTCGGCAG 3835  
 909 LeuSerLeuGlnAspAsnLeuGlnHisMetLeuSerProProGlnIleTh 925  
 3836 CTCTCACTTCAGGACACCTACAGCACATGCTCTCCCCCGCCCATCAC 3885  
 925 rIleGlyProGlnArgProAlaProSerGlyProGlyGlyGlySerGlyG 942  
 3886 CATCGGTCCCCAGAGGCCAGCTCCCTCAGGGCCAGGAGGGGGAGTGGTG 3935  
 942 lyGlySerGlyGlyGlyGlyGlyGlyGlnProProProProLeuGlnArgGly 958  
 3936 GGGGAGTGGTGGGGGGGGTGGGGGGCCAGCCACCTCCCTTGCAGAGGGGC 3985  
 959 LysSerGlnGlnLeuThrValSerAlaAlaGlnLysProArgProSerSe 975  
 3986 AAATCTCAGCAGTTGACAGTGTGCTGCCAGAAACCCCGCGCTCCAG 4035  
 975 rGlyAsnLeuLeuGlnSerProGluProSerTyrGlyProAlaArgProA 992  
 4036 CGGGAACCTATTGCAGTCCCCCGGAACCAAGTATGTTCTCTGCCGTCCAC 4085  
 992 rgGlnGlnSerLeuSerLysGluGlySerIleGlyGlySerGlyGlySer 1008  
 4086 GGCACAGAGCCTCAGCAAGAGGGCAGCATGGGGGCGAGCGGGGGCAGC 4135  
 1009 GlyGlyGlyGlyGlyGlyGlyLeuLysProSerIleThrLysGlnHisSe 1025  
 4136 GGTGGGGAGGGGGTGGGGGGCTCAAGCCCTCCATCACCAAGCAGCATTC 4185  
 1025 fGlnThrProSerThrLeuAsnProThrMetProAlaSerGluArgThrV 1042

4186 CCAGACTCCATCCAGCTGAACCCACGATGCCGGCTCGAGCGGACTG 4235  
 1042 alAlaTrpValSerAsnMetProHisLeuSerAlaAspIleGluSerAla 1058  
 4236 TAGCCTGGGTGTCCAATATGCTTCACTTCCGCTGACATCGAGAGTGCA 4285  
 1059 HisIleGluArgGluGluTyrLysLeuLysGluTyrSerLysSerMetAs 1075  
 4286 CACATTGAGCGGGAGAGTACAAGCTGAAGAGTACTCGAAGTCCATGGA 4335  
 1075 pGluSerArgLeuAspArgValLysGluTyrGluGluGluIleHisSerL 1092  
 4336 CGAGAGCCGACTGGACAGG.....GAGTACGAGGAGGAGTCCACTCAC 4379  
 1092 euLysGluArgLeuHisMetSerAsnArgLysLeuGluGluTyrGluArg 1108  
 4380 TGAAGGAAAGGCTACACATGTCCAACCCGAAGCTGGAAGAGTACGAGCGG 4429  
 1109 ArgLeuLeuSerGlnGluGluGlnThrSerLysIleLeuMetGlnTyrG1 1125  
 4430 AGGCTGCTGTCCCGAAGAGCAGACCCAGCAAGATCCTGTATCAGTACCA 4479  
 1125 nAlaArgLeuGluGlnSerGluLysArgLeuArgGlnGlnValGluL 1142  
 4480 AGCCCGCTTGAGCAGAGCGAGAGCGCTGAGCGCAGCAGCAGGTGGAGA 4529  
 1142 ysAspSerGlnIleLysSerIleIleGly.....Arg.LeuMe 1154  
 4530 AGGACTCCAGATCAAGAGCATCATTTGGCAGCCCGTCCCTTCAGGCTGAT 4579  
 1154 tLeuValGluGluGluLeuArgArgAspHisProAlaMetAlaGluProL 1171  
 4580 GCTGGTGGAGGAGGAGTGCCTGGGAGCACCCTCCGCTGAGCGCGC 4629  
 1171 euProGluProLysLysArgLeuLeuAspAlaGln.ArgGlySerPhePr 1187  
 4630 TGCTGAAACCCCAAGAGAGGCTGCTCGACGCTCAGGAGAGGAGCTTCCC 4679  
 1187 oProTrpValGlnGlnThrArgVal...ArgTrpProHisLeuGlyThrA 1203  
 4680 CCCTGGGTCCCAACAAACCCCGCTGTGACGCTGGCCCGCCACCTTGAACGG 4729  
 1203 laTrpProProGlnProHisProHisProGlyCysArgSerGlnArg 1219  
 4730 CCTGGCCCCCGCAGCCACCCCGCCCGCTGCGGCTGCAGATCACAGAGA 4779  
 1220 ThrAlaSerSerGlyThrPro 1226  
 4780 ACGGCAGTTCCGGAACACCG 4800

seq\_name: gb\_htg20:AL161903

seq\_documentation\_block:

LOCUS AL161903 150956 bp DNA HTG 29-SEP-2000  
 DEFINITION Homo sapiens chromosome 6 clone RP11-175A4, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 3 unordered pieces.

ACCESSION AL161903

VERSION AL161903.12 GI:10039702

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 150956)

AUTHORS Tracey,A.

TITLE Direct Submission

JOURNAL Submitted (28-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Sep 9, 2000 this sequence version replaced gi:9796876.

----- Genome Center

Center: Sanger Centre



106753 TAAGGTCCTTTGGCACATAAATCTATGAGTCTGTTCTTCCCAAACTT 106704  
 141 ..... 141  
 106703 CATGCTTCCAGTTCTGTGACGATATCTCTGTATCCAGTTTCTAGCAAGC 106654  
 141 ..... 141  
 106653 TTACAGTCCCTAGTCACACTCCTCTGGGGGAACTCTTGTGCTTGATGTC 106604  
 141 ..... 141  
 106603 CTAAAGAACTCAACCCCAAGATGCTGACATGATCTGACCTATGCAGAGTA 106554  
 141 ..... 141  
 106553 CAAACACCATGTTCCCTTTCAACACACTGCAATATCACCTGAGCCTTACC 106504  
 141 ..... 141  
 106503 ATTAGCTGTGCTCAATGTATGATCCCTCCCGGCTTACTGATCTTTTT 106454  
 141 ..... 141  
 106453 TTTTTTTTTTTGAGACAGAGTCTCACTCTGTCAACCCAGGCTGGAGTGCAG 106404  
 141 ..... 141  
 106403 TGGCGTGATCTTGGCTCACTGCAACCTCCGCCCTCTGGGTTTAAAGCAATT 106354  
 141 ..... 141  
 106353 CCTGTGCTCAGCTTCCCGAGTAGCTGGGATTACAGGCATGCACCACCAC 106304  
 141 ..... 141  
 106303 ACCCGACTAAATTTTGTATTTTAGTAGAGCAGGTTTCACCATGTTGG 106254  
 141 ..... 141  
 106253 CCAGGCTAGTCTCAAACCTCCTGACCTCAGTTGTCTGCTGCCCTCAGCCT 106204  
 141 ..... 141  
 106203 CCCAAAGCTAGGATTACAGGCATGAGCCACCATCTGGCCAGCTTT 106154  
 141 ..... 141  
 106153 ACCCATCTATATGCAATCTTTTTCACAAATTTCTGATAGTCTCTGCAG 106104  
 141 ..... 141  
 106103 GACTTTCCAGTCCCTTCAAATCTACATTAAATATTTTGGCTTGTATT 106054  
 141 ..... 141  
 106053 CCAGCTTTGAAATCTTTCCAATACTGTTTGTGTGTCTAGTGTATGCT 106004  
 141 ..... 141  
 106003 ACGCTTCTACAGTAGCTTCTGGGGCTGCTATATAACAACTGCTACAACT 105954  
 141 ..... 141  
 105953 GGGTCAGTAAACAAACAGAAATTTATTTCTCTACAGTTCAGGAGGCTGG 105904  
 141 ..... 141  
 105903 AAGTCCAAAAGCAAGGTATCAGCAGGGCCCGCTCTCTTTGATGCTGTAG 105854  
 141 , ..... 141

105853 CAGGGAATCCTTCCTTGCTCTTCTAGCTTCCGGAGGTTGCCAGCAGTC 105804  
 141 ..... 141  
 105803 CTTGGCATTCCTGGGCTTATAAAGTGCATCCCTCTAATTTCTGCCTCCATC 105754  
 141 ..... 141  
 105753 TTCATGTAGCTGGCTTCCTTCTGTGTCTCTGTATCCTGTATCTCTGTG 105704  
 141 ..... 141  
 105703 TCTCCAAATCTCCCTCTCCATATAAAGACACACAGTTTATTAAGGTGGT 105654  
 141 ..... 141  
 105653 TAAGGTCCTCACTCTAATTCAGTATGGCTCTTTTGTGATGGAATAT 105604  
 141 ..... 141  
 105603 CGCTCTTGTTCCTCCAGGCTTGGAGTGCAATGGCATGAGCTCAGCTCACTG 105554  
 141 ..... 141  
 105553 CAACCTCCGCCCTTCCAGGTTCAAGTGATTCTCTGCCTCAGCCTCCTGAG 105504  
 141 ..... 141  
 105503 TAGCTGGGATTACAGGCACGTCCTCACTATGCCAGCTAATTTGTTTTGT 105454  
 141 ..... 141  
 105453 ATTTTATTAGAGACGGGGTTTCACTATTTGGCCAGGCTGGCTCGAAT 105404  
 141 ..... 141  
 105403 CCTGACCTCAGGCAATCCACCCACCTCGGCTCCCAAAGTCTGGGATTA 105354  
 141 ..... 141  
 105353 CAGGAATGAGATACCATGGCTGGCTCTTCTTAGTTTGTATTACATTGCA 105304  
 141 ..... 141  
 105303 GGGATCCTGTTTCCAAATAAGATCACATTACAGGTTTCAGGGTAGACATG 105254  
 141 ..... 141  
 105253 AGTTTGGTGGGATACTATTCAACCCAGTACACCGTCTCACCTTATGATT 105204  
 141 ..... 141  
 105203 GCTGTCTGTATCCCACTTTCTTACTGTCTCTCTCTTTTGGTGTCTCT 105154  
 142 ..... AspAsnSe 144  
 105153 GTCCTTCCCTTTAAACATCATGCCACCCACCATGCCAGGACAACAG 105104  
 144 rArgArgValAspAsnValLeuLysLeuTrpIleIleGluAlaArgGluL 161  
 |||||  
 105103 CCGCGGGTAGACAATGTCTAAAGCTGTGATCATAGAGCCCGGGAGC 105054  
 161 euProProLysLysArgTyrTyrCysGluLeuCysLeuAspAspMetLeu 177  
 |||||  
 105053 TGCCCCCAAGAAGCGGTACTACTGTGAGCTCTGCCTGGATGACATGCTG 105004  
 178 TyrAlaArgThrThrSerLysProArgSerAlaSerGlyAspThrValph 194  
 |||||  
 105003 TATGCACGCACCATCTCAAGCCCGCTCTGCTCTGGGGACACCGCTCTT 104954  
 194 eTrpGlyGluHisPheGluPheAsnAsnLeuProAlaValArgAlaLeuA 211  
 |||||  
 104953 CTGGGGCGGAGCACACTCGAGTTTAAACAACCTGCCGGCTGTCCGTGCCCTGC 104904



211 rgLeuHisLeuTyrArgAspSerAspLysLysArgLysLysLysAla 227  
 104903 GGCTGCATCTGTACCGTGACTCAGACAAAAGGCGCAAGAGGCA 104854  
 228 GlyTyrValGlyLeuValThrValProValAlaThrLeuAlaGlyArgHi 244  
 104853 GGCTATGTCGGCCTGGTGACTGTGCCAGTGGCCACCCTGGCTGGCGCCA 104804  
 244 sPheThrGluGlnTrpTyrProValThrLeuProThrGlySerGlyGlys 261  
 104803 CTTTCACAGACAGTGGTACCTGTAAACCTGCCAACAGGCAGTGGGGAT 104754  
 261 erGlyGlyMetGlySerGlyGlyGlyGlyGlySerGlyGlySerGly 277  
 104753 CTGGGGGATGGGTTTCGGGGGGGGAGGGGGCTCGGGGGTGGCTCAGGG 104704  
 278 GlyLysGlyLysGlyGlyCysProAlaValArgLeuLysAlaArgTyrGl 294  
 104703 GGCAAGGGCAAGAGGTTGCCCGGCTGTGCCGCTGAAAGCACCGTTACCA 104654  
 294 nThrMetSerIleLeuProMetGluLeuTyrLysGluPheAlaGluTyrV 311  
 104653 GACAATGACCATCTTGCCCATGGAGCTATATAAAGAGTTTGCAGAGTATG 104604  
 311 alThrAsnHisTyrArgMetLeuCysAlaValLeuGluProAlaLeuAsn 327  
 104603 TCACCAACCATATTCCGGATGCTGTGCGAGTCTGGAGCCCGCCCTGAAT 104554  
 328 valLysGlyLysGluGluValAlaSerAlaLeuValHisIleLeuGlnSe 344  
 104553 GTCAAAGGCAAGGAGGAGTTCAGTGCACACTAGTTCACATCCTGCAGAG 104504  
 344 rThrGlyLysAla.Lys ..... 349  
 104503 TACAGGCAAGGCCAAGGTGAGTGTGTGCCCTCAGGGAAGGTGACTGG 104454  
 349 ..... 349  
 104453 GAATGGGCACCTGCTTGGGGGTTAGTGAGGACAGGCAAAATTCACGAGAT 104404  
 350 ..... AspPheL 352  
 104403 TGGGTTGTGCAGAGGCTGACACTTGGATTTTCTCGGCCCTCAGGACTCC 104354  
 352 euSerAspMetAlaMetSerGluValAlaAspArgPheMetGluArgGluHis 368  
 104353 TTTTCAGACATGGCCATGTCTGAGGTAGACCGGTTTCATGGAACGGGGAC 104304  
 369 LeuIlePheArgGluAsnThrLeuAlaThrLysAlaIleGluGluTyrMe 385  
 104303 CTCATATTCCGCGAGAACACGCTTGCCACTAAAGCCATAGAGAGTATAT 104254  
 385 tArgLeuIleGlyGlnLysTyrLeuLysAspAla.Ile..... 397  
 104253 GAGACTGATTGGTCAGAAATACCTCAAGGATGCCATGGTATGGCCACA 104204  
 397 ..... 397  
 104203 CTCAGGCCCTCTTCTTCCCAACCTGCCAGATGTCCACCCCGAGACCCCAA 104154  
 397 ..... 397  
 104153 GTCCACCCCTTCCACAGCTTGATACTTCTTAACCCAGAGTCTTAGGACTCC 104104  
 397 ..... 397  
 104103 AGCCTCCAACACCTGATTCTGAAATTTCCCAACCCCTGGCCACCCCTTC 104054  
 398 .....G1 398  
 104053 CCTGCCCCTTGGAAAGTGTGACCACACCCCTCTTGTGCCCCCACCACCCAGG 104004

398 yGluPheIleArgAlaLeuTyrGluSerGluGluAsnCysGluValAspP 415  
 104003 AGAATTATCCGTGCTCTGTATGAATCTGAGGAAAACCTGCGAGGTAGACC 103954  
 415 roIleLysCysThrAlaSerSerLeuAlaGluHisGlnAlaAsnLeuArg 431  
 103953 CTATCAAGTGCACAGCATCCAGTTTGGCAGAGCACCAGGCCAACCTGCGA 103904  
 432 MetCysCysGluLeuAlaLeuCysLysValValAsn..... 443  
 103903 ATGTGCTGTGAGTTGGCCCTGTGCAAGGTGGTCAACTCCCACTGGTGAGA 103854  
 443 ..... 443  
 103853 CTGGGAACGCTGGGCTGGGGGCCAGGGTCGGGGAATTATGTTCATC 103804  
 443 ..... 443  
 103803 TGTTCATCTATCTGTCCATCTCAAAAGAGGACTGAGCACCATTTATGGGC 103754  
 443 ..... 443  
 103753 AAAGCATTTCTTAGGCGCTATAGAGCAAAACAGGTGAAAAGAGGCTTGGTC 103704  
 443 ..... 443  
 103703 CCTGCCCTCAGAGGGCTCCACCAGAAATGGTGTAAAGTGTGAGFAAGGTCC 103604  
 443 ..... 443  
 103603 CCCCAACTCTGTGTGACATAAAGTGTGAGAGAGGAGGCTTTGAGAT 103554  
 443 ..... 443  
 103553 AAGTGGGAAGAGGTGCCCTTGGGTAGGCTTTGAAGACTGGTTAGGT 103504  
 443 ..... 443  
 103503 TCTGATATATGGACATAGTTGGCAAGAAAGACATTTTCAGAAAGGCTGT 103454  
 443 ..... 443  
 103453 GAGAAAGGCACATGTGTGATGGTGAAGAGGCCAGGAGTTTTCAGGGGAC 103404  
 443 ..... 443  
 103403 ATTAAAGTAGTTAGTAGCAATATACATCAGGTTTAGTGGAGCATGTGCCT 103354  
 443 ..... 443  
 103353 CATATGGGGAGTGGCGGGAGAGATGTCTGGCAGGAAGATTAGCTTTAG 103304  
 443 ..... 443  
 103303 AAAGTGAAGGCCCTCAAGGAGTCTGAGGTCAATGGTAGGCTTGGGATGC 103254  
 443 ..... 443  
 103253 CATTAAAGGTGTCAAGAAATATGTGATTTAGAAGATTAAATCCATAGGCTG 103204  
 444 .....SerHis.Cys..... 446  
 103203 GGCACGGTGGCTCACACCTGTAATACCAGCACTTTGGGAGGCTGAGGCGG 103154  
 446 ..... 446  
 103153 GCAGACCACCTGAGGTGAGGAGTTTGAGACCAGGCTGACCAACATGGAGA 103104  
 446 ..... 446

102203	C C C T C C T T C C C C T G T C T C T C T C A C C C C T G T T T C C A C A C C C T C A C C T C C T A C	102154
446	. . . . .	446
102153	C A C C C C C C T C A G C A T G T T C C C T G G A A G C T G A G G G T C T C T G G G G C T C A G T C	102104
446	. . . . .	446
102103	C C G G T C T C T C T T T C T C T C T C T C T C T C T C T C T C T C C C C G A C C C T T C C C C	102054
447	. . . . . ValPheProArgGluLeuLysGluValPheAlaSerTrpArgLeu	461
102053	C C A G C G T G T T C C C G A G G A G C T G A A G A G G T G T T T G C T T C A T G C G G C T G	102004
462	A r g C y s A l a G l u A r g G l y A r g G l u A s p I l e A l a A s p A r g L e u I l e S e r A l	478
102003	C G C T G C G C A G A G C G A G G C C G G A G G A C A T C G C A G A C A G S C T T A T C A G C G C	101954
478	a s e r L e u P h e L e u A r g P h e L e u C y s P r o A l a I l e M e t S e r P r o S e r L e u P	495
101953	C T C A C T C T T C C T G C G C T T C C T C T G C C A G C G A T T A T G T C G C C C A G T C T C T	101904
495	h e g l y L e u M e t G l n G l u T y r P r o A s p G l u G l n T h r S e r A r g T h r L e u T h r	511
101903	T T G G G C T T A T G C A G G A G T A C C C A G A T G A G C A G A C C T C A C G A A C C C T C A C C	101854
512	L e u I l e A l a L y s V a l I l e G l n A s n L e u A l a A s n P h e S e r L y s . . . . .	525
101853	C T C A T T G C C A A G T C A T C C A G A A C C T G C C A A C T T T T C C A A G T G A G G G A A	101804
525	. . . . .	525
101803	G C T T C A G G A T G G C C A G G C A G G A G T G C C A G G G C A G G A G T G C A G G G C	101754
525	. . . . .	525
101753	T G G G G T C G G C A A G A G G G T C T C C T G A G T C C C C A G A G A T C C T G A G A T G G G	101704
525	. . . . .	525
101703	G A G G C T A T G A T A C C T T G T G T G T G T A T G T G T G T G T G T G T G T G T G T G T	101654
526	. . . . . P	526
101653	G T G T G T G T G T G T A T G T G A C C T T A T C T T C T G C A T T C T T G G C T A G G T	101604
526	h e t h r S e r L y s G l u A s p P h e L e u G l y P h e M e t A s n G l u P h e L e u G l u L e u	542
101603	T T A C C T C A A A G G A G G A C T T T C T G G G C T T C A T G A A T G A G T T C T G G A G C T G	101554
543	G l u t r p G l y S e r M e t G l n G l n P h e L e u T y r G l u I l e S e r A s n L e u A s p T h	559
101553	G A A T G G G G T T C C A T G C A G C A G T T T T T G T A T G A G A T C T C C A A T C T G G A C A C	101504
559	r L e u T h r A s n S e r S e r P h e G l u T y r I l e A s p L e u G l y A r g G l u L	576
101503	G C T A A C C A A C A C A G C A G T A G C T T T G A G G G T T A C A T C G A C T T G G G C C G A G A G C	101454
576	e u S e r T h r L e u H i s A l a L e u L e u t r p l u V a l L e u P r o G l n L e u S e r L y s	592
101453	T C T C C A C A C T G C A T G C C C T A C T C T G G A G G T G C T G C C C A G C T C A G C A A G	101404
593	. . . . . Glu . . . . .	593
101403	G T C A G C A G A T C C C C T C T T T G C C C T A T C C C C A G A T G G C T C C A G A G G T T C C T	101354
593	. . . . .	593
101353	G G A G C C T G A G A A A C T A C C C T T T G A A G A T T T T T T T T C T C C C C T T G T T T C T C	101304
593	. . . . .	593
101303	G A G G T G T C A C C A C T A C T A T C C C A A C T C A G G C C C C C T C C A C C T G C A C C C T C	101254

593 ..... 593  
101253 AGAGGCCCTCTAGAGCTGGGCACTGAGCCCCCAGGTAACAGCCTACCC 101204  
594 .....AlaLeuLeuLysLeuGlyProLeuProArgLeuLeuSerAs 607  
101203 TTCCAGGAAGCCCTCCTGAAGCTGGGTCCACTGCCCGGCTCCTCAACGA 101154  
607 pIleSerThrAlaLeuArgAsnProAsnIleGlnArgGlnProSerArgG 624  
101153 CATCAGCACAGCTCTGAGGAACCCCAACATCCAAAGGCAGCCAGCCGCC 101104  
624 lnSerGluArgAlaArgSerGlnProMetValLeuArgGlyProSerAla 640  
101103 AGAGTGAGGCGCCCGCCCTCAGCCTGTGGTACTGCGGGGGCCATCGGCT 101054  
641 GluMetGlnGlyTyrMetMetArgAspLeuAsn ..... 651  
101053 GAGATGCAGGGCTACATGATGCGGGGACCTCAACAGGTGAGCACCCCTGGGA 101004  
651 ..... 651  
101003 CAGCCAGGCCCTGTGCCCTAGGAGCCCTTCTCCTATTCTAGATACTCCTCA 100954  
651 ..... 651  
100953 CTGGGCCCCACATGCATCTCTTAGGGCTTGAAAGAGGGAGGAAAAAGC 100904  
651 ..... 651  
100903 ACCAAGTTCTCAGGGGAGACGATAAGGAGACAGGTACAGTCAGTGGTAGG 100854  
651 ..... 651  
100853 CTGAGAGCCCTTTACAGCCTGAGGGAGTGAGAGATGTGGAGCTCTAGGAA 100804  
651 ..... 651  
100803 TAGGGCTGAGGCTCCACCAACTCAGGGCTTAGTTGTAAGCCTAGAGCATC 100754  
651 ..... 651  
100753 CCTGCTGCAAGCTCTGATTTGCTGTCCCTCTGCCTGCCCATGTAGTCCC 100704  
651 ..... 651  
100703 CAGGCTGAGGTTACGCCAGCATGTATGTACAGCCATGTGTCAAAATGTTT 100654  
651 ..... 651  
100653 AAACATCTCAGTAATAGCTAGTGAATAAGCACTTCCCCCAGCCCCCGACC 100604  
651 ..... 651  
100603 ACAACCCACAGACCTCCCCATGATCCAGCACTTAGAGCAGTGACAGCAG 100554  
651 ..... 651  
100553 AAGCCTAGCAGGCGCTGCAGCCTGCTCCAGAGTCCAGCCTCCATTCTGA 100504  
651 ..... 651  
100503 TAGGTGGTGCCCGTGCTTCAATTGCTGCTCATTTATTTGATGATGGCCCT 100454  
651 ..... 651  
100453 GCTTTTCTGCCCCGTGTCTTCTCCTCCATGACACCATACCCATCCACC 100404  
651 ..... 651  
100403 ATTCCCGCCTCTCCTTTCAATTGTCCACATCTCTCTCCTCTCTGTCTGT 100354

652 .....SerSerIleAspLeuGlnSerPhe 659  
100353 GCTCGCCCTCTTTTCCATCTCTTCCAGCTCCACCTTCAGTCTCCTTC 100304  
660 MetAlaArgGlyLeuAsn ..... 665  
100303 ATGGCTCGAGGCTCAACAGGTGAGGGGCTCTCCCTCCCGCCCTCCT 100254  
665 ..... 665  
100253 CTCTCTCTCTGTATGTTCCCTCTCCACCTCCACTGGCCTTCGCCCTACTC 100204  
665 ..... 665  
100203 CTCTCTCTCTCTCTCCATGGACCTCATCTCTCTCCATATGTGCCAGCCC 100154  
665 ..... 665  
100153 TGCCCCCATCCCTTCTCTGTGCTGCCCCCATCTCCCTCTCTTAGGCCCTCA 100104  
665 ..... 665  
100103 CCCCCTTCCCGGAGGGGCTCTCTCTTTCCCTTTACTCACCTGTGCCCTC 100054  
665 ..... 665  
100053 CCATCTCTCTCTGCTGCTTTCAGGGCTGCCACCGCTAGCTCTCAGCC 100004  
665 ..... 665  
100003 CTTCCCTCTGGGTCCCACCTTTTACCCCCAAGSCCTGTGCCAGACACAGC 99954  
665 ..... 665  
99953 AAGGTTCAATTGCTAGGAGCCCTGACCTTACCTTCTGTTGTGGCCCCC 99904  
666 .....Ser.SerMetAspMetAlaArgLeuProSerProThrLy 678  
99903 TTCCCTTCTGACAGCTCTATGGACATGGCTGGCTCCCTCCCAACCAA 99854  
678 sGluLysProProProProGlyGlyLysAspLeuPheTyrV 695  
99853 GGAAAGCCACCCCAACCGGCTGGTGGTAAAGACCTGTGTCTATG 99804  
695 alSerArgProProLeuAlaArgSerSerProAlaTyrCysThrSerSer 711  
99803 TAAGCGCTCCACCCCTGGCCGCTTCTCACCAGCATACTGCACGAGCAGC 99754  
712 SerAspIleThrGluProGluGlnLysMetLeuSerValAsnLysSerVa 728  
99753 TCGGACATCACAGAGCCAGACAGAGATGCTGAGTGTCAACAAGAGTGT 99704  
728 lSerMetLeuAspLeuGlnGlyAspGlyProGlyArgLeuAsnSerS 745  
99703 GTCCATGCTGGACTTACAGGGTGTGGGCTGGGCTGGGCTGGGCTCAACAGCA 99654  
745 erSerValSerAsnLeuAlaAlaValGlyAspLeuLeuHisSerSerGln 761  
99653 GCAGTGTTCGAACCTGGCGGCTGAGGGACCTGTGCTGCTCAAGCCAG 99604  
762 AlaSerLeuThrAlaAlaLeuGlyLeuArgProAlaProAlaArgLe 778  
99603 GCCTCGCTGACAGCAGCCTTGGGGCTACGGGCTGGGCTGGGCTGGGCT 99554  
778 uSerGlnGlySerGlySerSerIleThrAlaAlaGlyMetArgLeuSerG 795  
99553 CTCCCAGGGAGTGGCTCATCTCATCAGCGGCTGGGCTGGGCTGGGCT 99504  
795 lnMetGlyValThrThrAspGlyValProAlaGlnGlnLeuArgIlePro 811  
99503 AGATGGTGTCCACCACAGAGGCTGTCCCTGCCAGCAACTGCCAATCCCC 99454  
812 LeuSerPheGlnAsnProLeuPheHisMetAlaAlaAspGlyProGlyPr 828



99453 CTCTCTCCAGAACCTCTCTCCACATGGCTGCTGATGGCCAGGTCC 99404  
 828 oProAlaGlyHisGlySerSerGlyHisGlyProProSerSerHisH 845  
 99403 CCCAGGGCCCATGGAGGGGGCTGGCCATGGCCACCTCTCTCCATC 99354  
 845 isHisHisHisHisHisHisHisHisHisHisHisHisHisHis 861  
 99353 ACCACCACCACCACCACCACCACCACCAGGTGGAGAGCCCTGGGAC 99304  
 862 ThrPheAlaProPheHisGlyTyrSerLysSerGluAspLeuSerThrG1 878  
 99303 ACCTTTGCCCATTCATGGCTATAGCAAGAGTGAGGACCTCTCTCCGG 99254  
 878 yValProLysProProAlaAlaSerIleLeuHisSerHisSerTyrSera 895  
 99253 GGTCCCCAAGCCCCCTGCTGCCCTCCATCTCATAGCCACAGCTACAGTG 99204  
 895 spGluPheGlyProSerGlyThrAspPheThrArgArgGlnLeuSerLeu 911  
 99203 ATGAGTTTGGACCTCTGGCACTGACTTACCCCTGCCAGATCCACATTTGCTC 99154  
 912 GlnAspAsnLeuGlnHisMetLeuSerProProGlnIleThrIleGlyPr 928  
 99153 CAGGACAACCTGCAGACATGCTGTCCCTCCCTCCAGATCCACATTTGCTC 99104  
 928 oGlnArgProAlaProSerGlyProGlyGlySerGlyGlySerG 945  
 99103 CCAGAGGCCAGCCCCCTCAGGGCCTGGAGGTGGAGCGGTGGGGGCGAGCG 99054  
 945 lyGlyGlyGlyGlyGlyGlnProProProLeuGlnArgGlyLysSerGln 961  
 99053 CTGGGGTGGCGGGGCGCAGCCGCTCCATTCAGAGGGGCAAGTCTCAG 99004  
 962 GlnLeuThrValSerAlaAlaGlnLysProArgProSerSerGlyAsnLe 978  
 99003 CAGTTGACAGTACAGCGCAGCCAGAAACCCCGCCATCCAGCGGAATCT 98954  
 978 uLeuGlnSerProGluProSerTyrGlyProAlaArgProArgGlnGlns 995  
 98953 ATTGCAGTCCCCAGACCAAGTTATGGCCCGCCCTCCACGCAACAGA 98904  
 995 erLeuSerLysGluGlySerIleGlySerGlySerGlySerGlyGly 1011  
 98903 GCCTCAGCAAGGAGGAGCAGCATTTGGGGCAGCGGGGCGCGGTGGCGGA 98854  
 1012 GlyGlyGlyGlyLeuLysProSerIleThrLys..... 1022  
 98853 GGGGTGGGGGCTGAAGCCCTCCATCACCAGCAGGTAGGTGAAGGCAG 98804  
 1022 ..... 1022  
 98803 GAGGAAGCGGGCTGGGTCAACACAGGAGGGAAGAGAGATGGGGGT 98754  
 1022 ..... 1022  
 98753 GGGGTGAACAGAGTCTGTGGCCTGAAGTTACAATCTTCTTGCCTCCTT 98704  
 1022 ..... 1022  
 98703 TTGGCCATTAAACAAATGTGTATAGAGGCGCTGTATGTGCATGTGCTA 98654  
 1022 ..... 1022  
 98653 TGCCAGGCACTAAGGATACGGCAGTGAACCCCTCTTAGCAGCTCTCATTTCCA 98604  
 1022 ..... 1022  
 98603 GAGGGGATTATCCATAAGTAGAATGGGGTGACTGGAATAGGGCCAGCT 98554  
 1022 ,..... 1022

98553 GGGGGCTTGGAGGATGAAGATCAGATCTCTGAATCAGAACAACAGAGGAA 98504  
 1022 ..... 1022  
 98503 ATTCTCTGGCATTCAACTCAGATCTCTGGCATTCAGAGGTGATAAAGGCT 98454  
 1022 ..... 1022  
 98453 ACAGCAGGGCAATAGGACTGGGACTGTCTCTCTTTGGTCTGTGCTTTG 98404  
 1022 ..... 1022  
 98403 CCTCTAAATGTACCTGCTTTTCCACACTTTCTTTCTCTGTCGCC 98354  
 1023 .....GlnHisSerGlnThrProSerThrLeuAsnProT 1034  
 98353 TCACTGTGCTTGTCCAGCATCTTCAGACACCATCCACATGAACCCCA 98304  
 1034 hrMetProAlaSerGluArgThrValAlaTrpValSerAsnMetProHis 1050  
 98303 CAATGCCAGCCTCTGAGCGGACAGTGGGCTGCTCCAAACATGCCTCAC 98254  
 1051 LeuSerAlaAspIleGluSerAlaHisIleGluArgGluGluTyrLysLe 1067  
 98253 CTGTGGCTGACATCGAGAGTGGCCACATCGAGCGGGAAGAGTACAAAGCT 98204  
 1067 uLysGluTyrSerLysSerMetAspGluSerArgLeuAspArgVallLys 1083  
 98203 CAAGGAGTACTCAAAATCGATGGATGAGAGCCGCGGTGGATAGGTACGA 98155

seq\_name: gb\_pr7:HSDJ570F3

seq\_documentation\_block:

LOCUS HSDJ570F3 94770 bp DNA PRI 17-DEC-1999  
 DEFINITION Human DNA sequence from clone RP4-570F3 on chromosome 6. Contains a gene similar to Rattus norvegicus synaptotagmin 1 (synaptotagmin-1) gene, the C10orf135, the C10orf135 (polypeptide from patented cDNA Em:E06811) gene, the PHF1 (PHD finger protein 1) gene, the KNSL2 (kinesin-like 2) gene, the gene for acyl-protein thioesterase, ESTs, STSS, GSSs and a CpG island, complete sequence.

ACCESSION AL050332  
 VERSION AL050332.15 GI:6010176  
 KEYWORDS HTG; C10orf135; CpG island; finger protein; GTPase-activating protein; kinesin-like; KNSL2; p135; PHD; PHF1; ribosomal protein L12; RPL12; SynGAP.

SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 94770)  
 AUTHORS Mashreghi-Mohammadi, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

COMMENT On Oct 4, 1999 this sequence version replaced gi:5870478.  
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

COMMENT

This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 The following abbreviations are used to associate primary accession



Em:AI050991 Em:AF034177 Em:W39055 Em:AA622278 Em:AA069440  
Em:AI087362 Em:AI168128 Em:AI340323 Em:AI028045  
Em:AA934024 Em:AI073391 Em:AA331045 Em:AA708235  
Em:AI342398 Em:AA056366 Em:AI582796 Em:AI702297  
Em:AI340198 Em:AI541387 Em:N34313 Em:AI567726 Em:AI364919  
Em:AI310946 Em:AA181976 Em:AI042256 Em:R81592 Em:T49588  
Em:AI929571 Em:D52353 Em:AA035663 Em:N48534 Em:AI373814  
Em:AA971499 Em:H25294 Em:AA703121 Em:AA446226 Em:AA379516  
Em:AI313425 Em:W070287 Em:AA937820 Em:H21228 Em:AA471384  
Em:N71594 Em:AA095718 Em:AA377858 Em:H25235 Em:H52756  
Em:D52710 Em:AA329903  
/evidence=not\_experimental  
/product="dJ570F3.2 (POLYPEPTIDE FROM PATENTED CDNA  
EMBL:E06811")



6470 TGCTAGTAAGGGTCCCCCAACCTCTGTGTGACATAAAGGTACAGAGAA 6421  
 445 ..... 445  
 6420 GGCAGAGCTTTGAGATAAAGTGGGAAGAGGTGCCCTTGGGTAGGCTTT 6371  
 445 ..... 445  
 6370 GAAGACTGGTTAGGTTCTGATATATGGACATAGTTGGCAAGAAAGACAT 6321  
 445 ..... 445  
 6320 TTCAGAAAGGCTGTGAGAAAGGCACATGTGTGATGGTGAAAGGCCCA 6271  
 445 ..... 445  
 6270 GGAGTTTCAGGGGACATTAAAGTAGGTTAGTAGCAATTACATCAGGTTT 6221  
 445 ..... 445  
 6220 AGTGGAGCATGTGCCCTCATAATGGGGAGTGGCGGGAGAGATGTCTGGGA 6171  
 445 ..... 445  
 6170 GGAAGATTAGCTTTAGAACTGGAAGGCCCTCAAGGAGTCTGAGGTCATTG 6121  
 445 ..... 445  
 6120 GTAGGCCCTGGGATGCCATTAAAGGTGTGAGAAATATTGTGATTAGAAG 6071  
 445 ..... 445  
 6070 ATTAATCCATAGGCTGGGCACGGTGGCTCACACCTGTAATACCAGCACTT 6021  
 445 ..... 445  
 6020 TGGGAGGCTAGGGCGGCAGACCACCTGAGGTCAGGAGTTTGAGACCAGG 5971  
 445 ..... 445  
 5970 CTGACCAACATGGAGAAACCTGTCTCTACTAAAAATACAAAATTAGCCA 5921  
 445 ..... 445  
 5920 GATGTGGTGACACATGCCCTGTAATCCCAGCTATTCGGGAGGCTAAGGCAG 5871  
 445 ..... 445  
 5870 GAGAACTCACTGAATCTGGGAGGTGGAGGTTCTAGTGAGCCGAGATCACA 5821  
 445 ..... 445  
 5820 CGATTGCACTTAGCCTGGGCAACAAGAGTGAAACTCCATCTCAAAAAA 5771  
 445 ..... 445  
 5770 AAAAAATTAAATCCATAAGTAGAGTGTATGTAGTAGAGTAGAGGGATGTATG 5721  
 445 ..... 445  
 5720 TTGGGGATGACTGGAATAGAGCCAGCTGGGGCTCGAGGCGATGAAAGT 5671  
 445 ..... 445  
 5670 CAGATCCTGAATCAGAACAGTGACAGAAGTTAGGAAGGAGCTACTGGAAG 5621  
 445 ..... 445  
 5620 GACCCCTCTCGGAAAGGATCAGTAGGAATTGTCAGCTTATTGGAAAGGAG 5571  
 445 ..... 445

5570 GGAGCATGTCTGTGGGGAGTCACGGATGACTCAAGAGGCCCATGAGGCTGG 5521  
 445 ..... 445  
 5520 TGGTTGGGAGACCCGGTGGCCCTCATTTGACAACCCAGAAAGTCAGGAGGAG 5471  
 445 ..... 445  
 5470 AGCCAGTTGGAGGTGTGGGGGCGATGGTGAGCTCTGTCTTTACACCGCCG 5421  
 445 ..... 445  
 5420 AGTTTAAGGTGTCAGTGGGACATTTGAAGTGAAGTGGGAGGTGCGGGGAG 5371  
 445 ..... 445  
 5370 TGAGCTCTGAGCCATTCCAGGGACTGGGGATCATGCTGGGCAACCTCCA 5321  
 445 ..... 445  
 5320 TCCCATTTCCCTGGAATCCAGAGAGATTGGGGGTCGAGCTCCCTGTA 5271  
 445 ..... 445  
 5270 CCTCAAGTGACCCCTCCATCTCTCTCCCATCTCTGTCTCTCCCTGGTGTCT 5221  
 445 ..... 445  
 5220 GTTTTCT 5171  
 445 ..... 445  
 5170 CTCCACAGTGTCT 5121  
 445 ..... 445  
 5120 CTAATCTGTGTCT 5071  
 445 ..... 445  
 5070 ATCTTGCT 5021  
 445 ..... 445  
 5020 CACCCTCACCT 4971  
 445 ..... 445  
 4970 TCTCTGGGGCTCAGTCCCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4921  
 446 ..... CysValPheProArgGluLeuLysGluValPh 456  
 4920 TCCCGGACCTTCCCCCGCGGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4871  
 456 eAlaSerTrpArgLeuArgCysAlaGluArgGlyArgGluAspIleAlaA 473  
 4870 TGCTTCGTGGCGCTGCGCTGCGCAGAGCGAGGCGCGGAGGACATCGCAG 4821  
 473 spArgLeuIleSerAlaSerLeuPheLeuArgPheLeuCysProAlaIle 489  
 4820 ACAGGCTTATCAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4771  
 490 MetSerProSerLeuPheGlyLeuMetGlnGluTyrProAspGluGlnPh 506  
 4770 ATGTCGCGCGCT 4721  
 506 rSerArgThrLeuThrLeuIleAlaLysValIleGlnAsnLeuAlaAsn 523  
 4720 CTCACGAACCTCTCACCCTCATTTGCCAAGGTCTCTCTCTCTCTCTCTCTCT 4671  
 523 heSerLys..... 525  
 4670 TTTCCAAGTGAGGGAAGCTTTCAGGAGTGGGCGAGGCGAGGAGTGGCAGGG 4621

525	.....	525
4620	CAGGAGTGGCAGGGCTGGGGGTGGCAAGAGGGTCTCCTGAGTCCCA	4571
525	.....	525
4570	GAGATCCTGAGATGGGAGGCTATGATACCTTGTGTGTGTATGTGT	4521
525	.....	525
4520	GTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATCTTCT	4471
526	.....PheThrSerLysGluAspPheLeuGlyPheMetAs	537
4470	GCATTCTTGGCTAGGTTTACCTCAAAGGAGGACTTCTGGGCTTCATGAA	4421
537	nGluPheLeuGluLeuTrpGlySerMetGlnGlnPheLeuTyrGluI	554
4420	TGAGTTCTGGAGCTGGAATGGGTTCATGCAGCAGTTTGTATGAGA	4371
554	leSerAsnLeuAspThrLeuThrAsnSerSerPheGluGlyTyrIle	570
4370	TCTCCAATCTGGACACGCTAACCAACAGCAGTAGCTTGTGAGGTTACATC	4321
571	AspLeuGlyArgGluLeuSerThrLeuHisAlaLeuLeuTrpGluValle	587
4320	GACTGGGCCGAGAGCTCTCCACACTGCATGCCCTACTCTGGAGGTGCT	4271
587	uProGlnLeuSerLys..Glu.....	593
4270	GCCCCAGCTCAGCAAGGTCAGCAGATCCCCCTTTGGCCCTATCCCCAGAT	4221
593	.....	593
4220	GGCTCCAGAGGTTCCCTGGAGCCTGAGAACTACCCTTTGAAGATTTT	4171
593	.....	593
4170	TCRCCCTTGTTCGAGGTGTCACCACTACTATCCCACTCAGGCCCC	4121
593	.....	593
4120	CTCCACCTGCACCCCTCAGAGGCCCTCTTAGAGCTGGGCACCTGAGCCCCA	4071
594	.....AlaLeuLeuLysLeuGlyProLeuP	602
4070	GGTAACAGCCTCACCCCTCCAGGAAGCCCTCTGAAGCTGGGTCCACTGC	4021
602	roArgLeuLeuSerAspIleSerThrAlaLeuArgAsnProAsnIleGln	618
4020	CCGGCTCCTCAACGACATCAGCACAGCTGTGAGGAACCCCAACATCCA	3971
619	ArgGlnProSerArgGlnSerGluArgAlaArgSerGlnProMetValle	635
3970	AGCAGCCAAAGCCGCAGAGTGAGCGGCCCGGCCCTCAGCCTGTGGTACT	3921
635	uArgGlyProSerAlaGluMetGlnGlyTyrMetMetArgAspLeuAsn.	651
3920	GCGGGGCCATCGGCTTGAGATGCAGGGCTACATGATGCGGGACCTCAACA	3871
651	.....	651
3870	GGTGAGCACCCCTGGGACAGCCAGGCCCTGTGCCCTAGGAGCCCTTCTCCTA	3821
651	.....	651
3820	TTCATAGATACTCCTCACTGGGCCCCACATGCATCTCTCTAGGGCTTGAAA	3771
651	.....	651
3770	GAAGGGAGGAAAAAGCACCAAGTTCTCAGGGGAGACGATAAGGAGACAGG	3721

2820 TGTGCCAGACACAGCAAGGTTCAATTGCTAGGAGCCCTGACCTTACCTT 2771  
666 .....Ser..SerMetAspMetAlaArgL 673  
2770 CTGCTGTGTGCCCTTCCCTTCTGACAGCTCTATGGACATGGCTCGCC 2721  
673 euProSerProThrLysGluLysProProProProProProGlyGly 689  
2720 TCCCTTCCCAACCAAGAAAGCCACCCACCAACCCGCTGGTGGT 2671  
690 LysAspLeuPheTyrValSerArgProProLeuAlaArgSerSerProAl 706  
2670 AAAGACCTGTTCTATGTAAGCCGTCCACCCCTGGCCCTTCTCTCACCAGC 2621  
706 aTyrCysThrSerSerSerAspIleThrGluProGluGlnLysMetLeuS 723  
2620 ATACTGCACGAGCAGCTCGGACATCACAGAGCCAGAGCAGAGATGCTGA 2571  
723 erValAsnLysSerValSerMetLeuAspLeuGlnGlyAspGlyProGly 739  
2570 GTGTCAACAAGAGTGTGCCATGCTGGACTTACAGGGTGATGGCCCTGGT 2521  
740 GlyArgLeuAsnSerSerSerValSerAsnLeuAlaAlaValGlyAspLe 756  
2520 GGCCGCTCAACAGCAGCAGTGTTCGAACCTGGCGGCCGTAGGGGACCT 2471  
756 uLeuHisSerSerGlnAlaSerLeuThrAlaAlaLeuGlyLeuArgProA 773  
2470 GCTGCACTCAAGCCAGGCTCGCTGACAGCAGCCTTGGGGCTACGGCCTG 2421  
773 laProAlaGlyArgLeuSerGlnGlySerGlySerSerIleThrAlaAla 789  
2420 CGCCTGCGGACGCTCTCCAGGGAGTGGCTCATCCATCAGCGGGCT 2371  
790 GlyMetArgLeuSerGlnMetGlyValThrThrAspGlyValProAlaG1 806  
2370 GGCATGCGCCTCAGCCAGATGGGTGTCCACACAGACGGTGTCCCTGCCCA 2321  
806 nGlnLeuArgIleProLeuSerPheGlnAsnProLeuPheHisMetAlaA 823  
2320 GCAACTGGAATCCCCCTCTCCTTCCAGAACCTCTCTTCCACATGGCTG 2271  
823 laAspGlyProGlyProProAlaGlyHisGlyGlySerSerGlyHisGly 839  
2270 CTGATGGCCAGGTCCCCAGCGGCCATGGAGGGGGCGGTGGCCATGGC 2221  
840 ProProSerSerHisHisHisHisHisHisHisHisHisHisHisHis 856  
2220 CCACCTTCTCTCCATCACCAACCAACCAACCAACCAACCAACCAACCA 2171  
856 yGluProProGlyAspThrPheAlaProPheHisGlyTyrSerLysSerG 873  
2170 AGAGCCCTGCGGACACCTTGGCCCATTCATCGCTATAGCAAGAGTG 2121  
873 luAspLeuSerThrGlyValProLysProProAlaAlaSerIleLeuHis 889  
2120 AGGACCTCTCTCCGGGGTCCCAAGCCCTGCTGCCTCCATCCTTCAT 2071  
890 SerHisSerTyrSerAspGluPheGlyProSerGlyThrAspPheThrAr 906  
2070 AGCCACAGCTACAGTATGATGTTGGACCCCTGCGCACTGACTCACCCG 2021  
906 gArgGlnLeuSerLeuGlnAspAsnLeuGlnHisMetLeuSerProProG 923  
2020 TCGGACGCTTCACTCCAGGAAACCTGCAGACATGCTGTCCCTCCCTCCC 1971  
923 lnIleThrLeGlyProGlnArgProAlaProSerGlyProGlyGlyGly 939  
1970 AGATCACCATGGTCCCAAGAGCCAGCCCTCAGGGCCTGAGGGTGGG 1921  
940 SerGlyGlyGlySerGlyGlyGlyGlyGlyGlyGlyGlnProProProLeuG1 956  
|||||

1920 AGCGGTGGGGGACGCGGTGGGGGTGGGGGGGCGCCAGCCGCTCCATTGCA 1871  
956 nArgGlyLysSerGlnGlnLeuThrValSerAlaAlaGlnLysProArgP 973  
1870 GAGGGCAAGTCTCAGCAGTTGACAGTCAAGCAGCCAGCCAGAAACCCCGG 1821  
973 roSerSerGlyAsnLeuLeuGlnSerProGluProSerTyrGlyProAla 989  
1820 CATCCAGCGGAATCTATTGAGTCCCCAGAGCAAGTTATGGCCCCGCC 1771  
990 ArgProArgGlnGlnSerLeuSerLysGluGlySerIleGlyGlySerG1 1006  
1770 CGTCCACGCAACAGAGCCTCAGCAAGGAGGGGAGCATTTGGGGGACGGG 1721  
1006 yGlySerGlyGlyGlyGlyGlyGlyGlyGlyLeuLysProSerIleThrLys. 1022  
1720 GGGCAGCGGTGGCGGAGGGGGTGGGGGCTGAAGCCCTCCATCACCAAGC 1671  
1022 ..... 1022  
1670 AGGTAGTGAAGGCAGGAGGAGGCGGGCTGGGTCAACAACAGGGAGGGAA 1621  
1022 ..... 1022  
1620 GAAGGAGATGGGGGTGGGGTTGAAACAGAGTCTGTGGCCTGAAGTTACA 1571  
1022 ..... 1022  
1570 ATCTTCTTGCCTCCTTTTGGCCATTAAACAAATGTGTATAGAGGGCCTGC 1521  
1022 ..... 1022  
1520 TATGTGCCATGTGCTATGCCAGGCACCTAAGGATACGGCAGTGAACCTCT 1471  
1022 ..... 1022  
1470 TAGCACTCTCATTTCCAGAGGGGATTAATCCATAAGTAGAATGGGGTGAC 1421  
1022 ..... 1022  
1420 TGAATAGGCCACAGCTGGGGGCTTGGAGGCATGAAGATCAGATCCTGAAT 1371  
1022 ..... 1022  
1370 CAGAACAACAGGAGAAATTTCTTGGCATTTCAACTCACATCTCTGGCATTC 1321  
1022 ..... 1022  
1320 AGAGGTATAAAGGCTACAGCAGGGCAATAGGACTGGGACTGTCGCTCC 1271  
1022 ..... 1022  
1270 TTTGGCTGTGCTGTGGCTCTFAAATGTACCCTGCTTTTCCACCTTTT 1221  
1023 .....GlnHisSerGlnThrPro 1028  
1220 CTTTCTCTGTTCGCCCTCAGTGTGCTTGTCCCAGCATTTCTCAGACACCA 1171  
1029 SerThrLeuAsnProThrMetProAlaSerGluArgThrValAlaTrpVa 1045  
1170 TCCACATTGAACCCCAATGCCAGCCTCTGAGCGGACAGTGGCCTGGGT 1121  
1045 lSerAsnMetProHisLeuSerAlaAspIleGluSerAlaHisIleGluA 1062  
1120 CTCCAACATGCTCCTGCTGCTGCTGACATCGAGAGTGGCCACATCGAGC 1071  
1062 rgGluGlyTyrLysLeuLysGlyTyrSerLysSerMetAspGluSerArg 1078  
1070 GGAAGAGTACAAGCTCAAGGAGTACTCAAAATCGATGGATGAGAGCCGG 1021  
1079 LeuAspArgValLys 1083  
1020 CTGGATAGGGTACGA 1006



seq\_name: gb\_htg20:AL161906

seq\_documentation\_block:

LOCUS AL161906 169323 bp DNA HTG 18-AUG-2000  
 DEFINITION Homo sapiens chromosome 5 clone RP11-567N9, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 24 unordered pieces.

ACCESSION AL161906  
 VERSION AL161906.5 GI:9863619  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 169323)

REFERENCE

AUTHORS  
 TITLE Direct Submission  
 JOURNAL Submitted (17-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Aug 21, 2000 this sequence version replaced gi:9212890.

COMMENT

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA567N9

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 155424 bases at least Q40

Consensus quality: 160003 bases at least Q30

Consensus quality: 163252 bases at least Q20

Insert size: 167023; sum-of-contigs

Insert size: 186175; 5.7% error; agarose-fp

Quality coverage: 3.59x in Q20 bases; sum-of-contigs Quality

coverage: 3.22x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 24 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 5456: contig of 5456 bp in length  
 \* 5457 5556: gap of 100 bp  
 \* 5557 8767: contig of 3211 bp in length  
 \* 8768 8867: gap of 100 bp  
 \* 8868 17168: contig of 8301 bp in length  
 \* 17169 17268: gap of 100 bp  
 \* 17269 26591: contig of 9323 bp in length  
 \* 26592 26691: gap of 100 bp  
 \* 26692 31188: contig of 4497 bp in length  
 \* 31189 31288: gap of 100 bp  
 \* 31289 37579: contig of 6291 bp in length  
 \* 37580 37679: gap of 100 bp  
 \* 37680 45350: contig of 7671 bp in length  
 \* 45351 45450: gap of 100 bp  
 \* 45451 48354: contig of 2904 bp in length  
 \* 48355 48454: gap of 100 bp  
 \* 48455 63401: contig of 14947 bp in length  
 \* 63402 63501: gap of 100 bp  
 \* 63502 67724: contig of 4223 bp in length  
 \* 67725 67824: gap of 100 bp  
 \* 67825 70550: contig of 2726 bp in length  
 \* 70551 70650: gap of 100 bp  
 \* 70651 80664: contig of 10014 bp in length  
 \* 80665 80764: gap of 100 bp  
 \* 80765 83307: contig of 2543 bp in length

\* 83308 83407: gap of 100 bp  
 \* 83408 88383: contig of 4976 bp in length  
 \* 88384 88483: gap of 100 bp  
 \* 88484 93697: contig of 5214 bp in length  
 \* 93698 93797: gap of 100 bp  
 \* 93798 99354: contig of 5557 bp in length  
 \* 99355 99454: gap of 100 bp  
 \* 99455 111055: contig of 11601 bp in length  
 \* 111056 111155: gap of 100 bp  
 \* 111156 118645: contig of 7490 bp in length  
 \* 118646 118745: gap of 100 bp  
 \* 118746 122421: contig of 3676 bp in length  
 \* 122422 122521: gap of 100 bp  
 \* 122522 132468: contig of 9947 bp in length  
 \* 132469 132568: gap of 100 bp  
 \* 132569 137959: contig of 5391 bp in length  
 \* 137960 138059: gap of 100 bp  
 \* 138060 151691: contig of 13632 bp in length  
 \* 151692 151791: gap of 100 bp  
 \* 151792 158253: contig of 6462 bp in length  
 \* 158254 158353: gap of 100 bp  
 \* 158354 169323: contig of 10970 bp in length.

FEATURES

Source

1. 169323  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="RP11-567N9"  
 /clone\_lib="RPC1-11.2"  
 1. 5456

misc\_feature

/note="assembly\_fragment:00884"

misc\_feature

fragment\_chain:1  
 5557. 8767  
 /note="assembly\_fragment:00832"

misc\_feature

fragment\_chain:1  
 8868. 17168  
 /note="assembly\_fragment:01060"

misc\_feature

fragment\_chain:1  
 17269. 26591  
 /note="assembly\_fragment:01172"

misc\_feature

fragment\_chain:1  
 26692. 31188  
 /note="assembly\_fragment:00279"

misc\_feature

fragment\_chain:1  
 31289. 37579  
 /note="assembly\_fragment:01205"

misc\_feature

fragment\_chain:1  
 37680. 45350  
 /note="assembly\_fragment:01334"

misc\_feature

fragment\_chain:2  
 45451. 48354  
 /note="assembly\_fragment:01263"

misc\_feature

fragment\_chain:2  
 48455. 63401  
 /note="assembly\_fragment:00990"

misc\_feature

fragment\_chain:2  
 63502. 67724  
 /note="assembly\_fragment:00045"

misc\_feature

fragment\_chain:2  
 67825. 70550  
 /note="assembly\_fragment:00965"

misc\_feature

fragment\_chain:2  
 70651. 80664  
 /note="assembly\_fragment:00637"

misc\_feature

fragment\_chain:3  
 80765. 83307  
 /note="assembly\_fragment:01291"

misc\_feature

fragment\_chain:3  
 83408. 88383  
 /note="assembly\_fragment:01317"

misc\_feature

fragment\_chain:3  
 88484. 93697  
 /note="assembly\_fragment:01171"

```

misc_feature      fragment_chain:4"
93798..99354
/note="assembly_fragment:00288
fragment_chain:4"
99455..111055
/note="assembly_fragment:00286
fragment_chain:4"
11156..118645
/note="assembly_fragment:00303
fragment_chain:5"
118746..122421
/note="assembly_fragment:01184
fragment_chain:5"
122522..132468
/note="assembly_fragment:00501
fragment_chain:6"
132569..137959
/note="assembly_fragment:01244
fragment_chain:6"
138060..151691
/note="assembly_fragment:00032"
151792..158253
/note="assembly_fragment:00779"
158354..169323
/note="assembly_fragment:01270"
BASE COUNT 41674 a 41332 c 41269 g 42673 t 2375 others
ORIGIN

alignment_scores:
  Quality: 3987.00      Length: 2457
  Ratio: 4.237          Gaps: 12
Percent Similarity: 38.299 Percent Identity: 37.810

alignment_block:
US-09-294-298-6 x AL161906 ..
Align seg 1/1 to: AL161906 from: 1 to: 169323

139 ProAsnLysAspAsnSerArgArgValAspAsnValLeuLysLeuTrpI1 155
||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
77990 CCATGCCAGGACACAGCCGCGGTAGACAATGTCTAAAGCTGTGGAT 78039

155 eIleGluAlaArgGluLeuProProLysLysArgTyrTyrCysGluLeuC 172
|||||
78040 CATAGAGCCCGGGAGCTGCCCCCAAGAGCGGTACTACTGTGAGCTCT 78089

172 ysLeuAspAspMetLeuTyrAlaArgThrThrSerLysProArgSerAla 188
|||||
78090 GCCTGGATGACATGCTGTATGCAGCGCACCCACCCTCCAAAGCCCGCTGCC 78139

189 SerGlyAspThrValPheTrpGlyGluHisPheGluPheAsnAsnLeuPr 205
|||||
78140 TCTGGGGACACCGTCTTCTGGGGCGAGCACCTTCGAGTTTAACAACCTGCC 78189

205 oAlaValArgAlaLeuArgLeuHisLeuTyrArgAspSerAspLysLysA 222
|||||
78190 GGCTGTCCGTGCCCTGCGGCTGCATCTGTACCGTGAATCAGACAAAAGC 78239

222 rgLysLysAspLysAlaGlyTyrValGlyLeuValThrValProValAla 238
|||||
78240 GCAAGAAGGACAAGGCAGGCTATGTCGGCTGTGACTGTGCCAGTGCC 78289

239 ThrLeuAlaGlyArgHisPheThrGluGlnTrpTyrProValThrLeuPr 255
|||||
78290 ACCCTGGCTGGGGCCCACTTCACAGAGCAGTGTACCTGTAACTGTGCC 78339

255 oThrGlySerGlyGlySerGlyMetGlySerGlyGlyGlyGlyGlyS 272
|||||
78340 AACAGGCACTGGGGGATCTGGGGCATGGGTTCGGGAGGGGGGGGGCT 78389

272 erGlyGlyGlySerGlyGlyLysGlyLysGlyGlyCysProAlaValArg 288
|||||
```

```

78390 CGGGGGGTGGCTCAGGGGGCAAGGGCAAGGAGGTGCCCCGGCTGTGCGG 78439
289 LeuLysAlaArgTyrGlnThrMetSerIleLeuProMetGluLeuTyrLy 305
|||||
78440 CTGAAAGCACGTTACCAGACAAATGAGCATCTTGCCCATGGAGCTATATAA 78489
305 sGluPheAlaGluTyrValThrAsnHisTyrArgMetLeuCysAlaValL 322
|||||
78490 AGAGTTTGCAGAGTATGTCACCAACCATTTATCGGATGCTGTGTGCAGTCT 78539
322 euGluProAlaLeuAsnValLysGlyLysGluGluValAlaLysAlaLeu 338
|||||
78540 TGGAGCCCGCCCTGAATGTCAAAGCAAGGAGGAGGTGCCAGTGCACATA 78589
339 ValHisIleLeuGlnSerThrGlyLysAla.Lys..... 349
|||||
78590 GTTCACATCCTGCAGAGTACAGGCAAGGCCAAGGTGAGTGTGTGTCCTC 78639
349 ..... 349
78640 AGGGAAGGTGACTTGGGAATGGGCACCTTGCTTGGGGGTTAGTGAGGACA 78689
349 ..... 349
78690 GGGCAAAATTCAGAGATTGGGTTGTGCAGAGGCTGACACTTGGATTTC 78739
350 .....AspPheLeuSerAspMetAlaMetSerGluValAspArgP 363
|||||
78740 TGGGCCTCAGGACTTCCTTTCAGACATGGCCATGTCTGAGGTAGACCGST 78789
363 heMetGluArgGluHisLeuIlePheArgGluAsnThrLeuAlaThrLys 379
|||||
78790 TCATGGAACGGGAGCACCTCATATTCGCGAGAACACGCTTGCCACTAAA 78839
380 AlaIleGluGluTyrMetArgLeuIleGlyGlnLysTyrLeuLysAspAl 396
|||||
78840 GCCATAGAAGAGTATATGAGACTGATGGTGGTCAGAAATACCTCAAGGATGC 78889
396 a.Ile..... 397
|:::
78890 CATGGTATGGCCACACTCAGGCCCTCTTCTTCCCAAAACCTGCCAGATG 78939
397 ..... 397
78940 TCCACCCCAAGCCCAAGTCCACCCTTCCACAGCTTGATACTTCCTAAC 78989
397 ..... 397
78990 CAGAGTCTTAGGACTCCAGCCTCCAAACACCTGATTCTGAAATTTCCCCAA 79039
397 ..... 397
79040 CCCTGGCCACCCCTTCCCTGCCCTTGGAAAGTGTGACACACCCCTCTTG 79089
398 .....GlyGluPheIleArgAlaLeuTyrGluSerGluG 409
|||||
79090 TGCCCCCACCCCCAGGAGAAATTCATCCGTGCTCTGTATGAATCTGAGGA 79139
409 uAsnCysGluValAspProIleLysCysThrAlaSerSerLeuAlaGluH 426
|||||
79140 AAAGTGGAGGTAGACCCCTATCAAGTGCACAGCATCCAGTTGGCAGAGC 79189
426 isGlnAlaAsnLeuArgMetCysGluLeuAlaLeuLysLysValVal 442
|||||
79190 ACCAGGCCAACCTGCGAATGTGCTGTGAGTTGGCCCTGTGCAAGGTGGT 79239
443 AsnSerHis..... 445
|||||
79240 AACTCCCACTGGTGAGACTGGGAACGCTGGGCTGGGGGGGCCAGGGTCGG 79289
445 ..... 445
79290 GGAATTATGTGTTTCATCTGTTCATCTATCTGTCCATCTCAAGAGGACT 79339
```

445 ..... 445  
79340 GAGCACCATTATGGGCAAGCATTTGTTCTAGGCGCTATAGAGCAACAG 79389  
445 ..... 445  
79390 GTGAAAGAGGCCTGGTCCCTGCCCTCAGAGGGCCTCCACCAGAAATGGGA 79439  
445 ..... 445  
79440 CAAATTAGAAGAAAAAAAAAAGCCACAGGCCATAATGGTGTGTAAG 79489  
445 ..... 445  
79490 TGCTGAGTAAGGGTCCCCCAACCTCTGTGTGACATAAGGTCAGAGAGAA 79539  
445 ..... 445  
79540 GGCAGAGCTTTGAGATAAGTGGGGAAGAGGTGCCCCCTTGGGTAGGCTTT 79589  
445 ..... 445  
79590 GAAGACTGGTTTAGGTTCTGATATATGGACATAGTTGGCAAGAAAGACAT 79639  
445 ..... 445  
79640 TTCAGAAGAAGGCTGTGAGAAAGGCACATGTGTGATGGTGAAGAGGCCCA 79689  
445 ..... 445  
79690 GGAGTTTCAGGGGACATTAAAGTAGGTTAGTAGCAATTACATCAGGTTT 79739  
445 ..... 445  
79740 AGTGGAGCATGTGCTCATAATGGGGAGTGGGGGAGAGATGTCTGGCA 79789  
445 ..... 445  
79790 GGAAGATTAGCTTTAGAAACTGGAAGGCCCTCAAGGAGCTGAGGTCATTG 79839  
445 ..... 445  
79840 TTAGGCCTTGGGATGCCATTAAAGGTGTGAGAAATATTGTGATTAGAAG 79889  
445 ..... 445  
79890 ATTAATCCATAGGCTGGGCACGGTGGCTCACACCTGTAAATACCAGCACTT 79939  
445 ..... 445  
79940 TGGAGGCTGAGCGGCGCAGACCACCTGAGGTGAGGATTGAGACCAGG 79989  
445 ..... 445  
79990 CTGACCAACATGGAGAAACCCCTGTCTCTACTAAAAATACAAAATTAGCCA 80039  
445 ..... 445  
80040 GATGTGGTGGCACATGCGCTGTAATCCCAGCTATTCGGGAGGCTAAGGCAG 80089  
445 ..... 445  
80090 GAGAACTCACTTGAATCTGGGAGGTGGAGGTTGTAGTGAGCCGAGATCACA 80139  
445 ..... 445  
80140 CCAATTGCACCTAGCCTGGGCAACAAGAGTGAAGAACTCCGTCTCAAAAAA 80189  
445 ..... 445  
80190 AAAAAATAATCCATAAGTAGAGTGTATGTAGTAGAGTAGAGGATGTATG 80239

445 ..... 445  
80240 TTGGGGGATGACTGGAATAGAGCCAGCTGGGGGCTCGGAGGCATGAAAGT 80289  
445 ..... 445  
80290 CAGATCCTGAATCAGAAACAGTGACAGAAAGTTAGGAAGGAGCTACTGGAAG 80339  
445 ..... 445  
80340 GACCCCTCTCGGGAAGGATCAGTAGGAATGTGAGCTTATTTGAAAGGAG 80389  
445 ..... 445  
80390 GGAGCATGTCTGTGGGAGTCACGGATGACTCAAGAGGCCATGAGGCTGG 80439  
445 ..... 445  
80440 TGGTTGGGAGACCCTGGCCCTCATTTGACAACCAAGAAAGTCAGGAGGAG 80489  
445 ..... 445  
80490 AGCCAGTTGGAGGTGTGGGGCGATGTGAGCTCTGCTTTACACCAGCCG 80539  
445 ..... 445  
80540 AGTTAAGGTGTGAGTGGGACATTTGAAGTGAAGTGGGAGTGGGGGAG 80589  
445 ..... 445  
80590 TGAGCTCTGAGCCATTCCAGGACTGGGGATCATGCCTGGGCACCTCCA 80639  
445 ..... 445  
80640 TCCCATTTCCCTGGATCCAGAGANNNNNNNNNNNNNNNNNNNNNNNN 80689  
445 ..... 445  
80690 NNN 80739  
445 ..... 445  
80740 NNN 80789  
445 ..... 445  
80790 GGTCCGAGCTCCCTGTACCTCAAGTGACCCCTCCATCTCTCTCCCATCTCT 80839  
445 ..... 445  
80840 GTCTCTCCCTGGTGTGTGTTTCTTCTCCTCCTCTCTCTCTCTCTCTCC 80889  
445 ..... 445  
80890 CACACCCCTCCATCTCTCTCCACAGTGTCTCTCCCTCACCTTCTCTCTCC 80939  
445 ..... 445  
80940 CCTCCATTCTCTCTCCCTAATCTGTGTGTTCTCCTCTCTCTCTCTCTCT 80989  
445 ..... 445  
80990 TCTTCAAGCAGCCCTCCCATCTTGTCTCTGGGTCCCTCTCTCTCTCTCTCT 81039  
445 ..... 445  
81040 TCTCACCCTGTTTCCACACCCCTCACCTCTCTCTCTCTCTCTCTCTCTCT 81089  
445 ..... 445  
81090 TCCCTGGAAGCTGAGGGTCTCTCTGGGGCTCAGTCCCGGCTCTCTCTCTCT 81139  
446 ..... CysvalpheProArg 450



82040	GAAGCTGGGTCCACTGCCCGGCTCCTCAACGACATCAGCACAGCTCTGA	82089
613	rgAsnProAsnIleGlnArgGlnProSerArgGlnSerGluArgAlaArg 	629
82090	GGAACCCCAACATCCAAAGGCAGCCAAAGCCGCGAGGTGAGCGGCCCGG	82139
630	SerGlnProMetValLeuArgGlyProSerAlaGluMetGlnGlyTyrMe      :::	646
82140	CCTCAGCCTGTGGTACTGCGGGGGCCATCGGCTGAGATGCAGGGCTACAT	82189
646	tMetArgAspLeuAsn. 	651
82190	GATGCGGGACCTCAACACAGGTGAGCACCCCTGGGACAGCCAGGCCTGTGCC	82239
651		651
82240	TAGGAGCCCTTCTCCTATTCTAGATACTCCTCACTGGGCCCCACATGCAT	82289
651		651
82290	CTCTCTAGGCGTTGAAAGAAGGGAGGAAAAAGCAAGTTCTCAGGGGA	82339
651		651
82340	GACGATAAGGAGACAGGTACAGTCAGTGGTAGGCTGAGAGCCCTTTACAG	82389
651		651
82390	CCTGAGGGAGTGAGAGATTGGAGCTCTAGGAATAGGCGTGAGGCTCCAC	82439
651		651
82440	CAACTCAGGCTTAGTTGTAAAGCCTAGAGCATCCCTGCTGCAAGCTCTGA	82489
651		651
82490	TTTGCTGTCCCTCTGCCTGCCCATGTAGTCCCCAGGCTGAGGTTCAGCC	82539
651		651
82540	AGCATGTCATGTCAGCCATGTGTCAAAATGTTCAAACATCTCAGTAATAG	82589
651		651
82590	CTAGTGAATAAGCACTTCCCCAGCCCCCGACCACACACCCACAGACCTC	82639
651		651
82640	CCCATGATCCAGCACTTAGAGCAGTGACAGCAGAAAGCCTAGCAGGGCCTG	82689
651		651
82690	CAGCCTGCTCCAGAGTCCCAGCCTCCATTTCTGATAGGTGGTGCCCGTGCT	82739
651		651
82740	TCATTGTGCTGCTCATTTATTGATGATGGCCCCCTGCTTTTCCCTCTGCCCG	82789
651		651
82790	TGCTTTCTCCATGACACCATACCCATCCCACCATTCCTCCGCTCTCCTTT	82839
651		651
82840	CATTGTCCACATCTCTCTCCTTCTGTCTGTGCTCGCCCCCTCTTTCCA	82889
652	.....SerSerIleAspLeuGlnSerPheMetAlaArgGlyLeuAs 	665
82890	TCTCTCTCCAGCTCCATCGACCTTCAGTCTTTCATGGCTCGAGGCCCTCAA	82939
665	n 	665
82940	CAGGTGAGGGGTCTCCCCCTCCCCCGCCCTCCTCTCTCTCTCTCTGTGT	82989

[illegible]

756	uLeuHisSerSerGlnAlaSerLeuThrAlaAlaLeuGlyLeuArgProA	773
83890	GCTGCACTCAAGCCAGGCCTCGCTGACAGCAGCCTTGGGGCTACGGCGCTG	839339
773	laProAlaGlyArgLeuSerGlnGlySerGlySerSerIleThrAlaAla	789
83940	CGCCTGCGGAGCCCTCTCCAGGGGAGTGCGTCCATCATCACGGCGGCT	83989
790	GlyMetArgLeuSerGlnMetGlyValThrThrAspGlyValProAlaG1	806
83990	GGCATGCGCCTCAGCCAGATGGGTGTACACACAGACGGTGTCCCTGCCCA	84039
806	nGlnLeuArgIleProLeuSerPheGlnAsnProLeuPheHisMeta1aA	823
84040	GCAACTGCGAATCCCCCTCTCCCTTCAGAAACCCTCTCTTCCACATGGCTG	84089
823	laAspGlyProGlyProProAlaGlyHisGlyGlySerSerGlyHisGly	839
84090	CTGATGGGCCAGGTCCCCCAGGCGGCATGGAGGGGCGGTGGCCATGGC	84139
840	ProProSerSerHisHisHisHisHisHisHisHisHisArgGlyG1	856
84140	CCACCTTCTCCCATCACCAACCACCAACCATCACCAACCACCGAGGTGG	84189
856	yGluProProGlyAspThrPheAlaProPheHisGlyTyrSerLysSerG	873
84190	AGAGCCCCCTGGGGACACCTTTGCCCCCATTCATGGCTATAGCAAGAGTG	84239
873	luAspLeuSerThrGlyValProLysProProAlaAlaSerIleLeuHis	889
84240	AGGACCTCTCTTCCGGGTCCCCAAGCCCCCTGTGCTGCCCTCCATCTCAT	84289
890	SerHisSerTyrSerAspGluPheGlyProSerGlyThrAspPheThrAr	906
84290	AGCCACAGCTACAGTGATGAGTTTGGACCCCTCTGGCACTGACTTCAACCCG	84339
906	gArgGlnLeuSerLeuGlnAspAsnLeuGlnHisMetLeuSerProProG	923
84340	TCGGCAGCTTTCACCTCCAGGACAACCTGCAGCACATGCTGTCCCCCTCCCC	84389
923	InIleThrIleGlyProGlnArgProAlaProSerGlyProGlyGlyGly	939
84390	AGATCACCATTTGGTCCCCAGAGGCCAGCCCCCTCAGGGCCTGGAGGTGGG	84439
940	SerGlyGlyGlySerGlyGlyGlyGlyGlyGlyGlnProProProLeuG1	956
84440	AGCGTGGGGGCAGCGGTGGGGTGGCGGGGGCCAGCCGCCTCCATTTGCA	84489
956	nArgGlyLysSerGlnGlnLeuThrValSerAlaAlaGlnLysProArgP	973
84490	GAGGGCAAGTCTCAGCAGTTGACAGTCAGCGCAGCCCCAGAAACCCCGGC	84539
973	roSerSerGlyAsnLeuLeuGlnSerProGluProSerTyrGlyProAla	989
84540	CATCCAGCGGGAAATCTATTGCAGTCCCCAGAGCCAAAGTTATGGCCCCCGCC	84589
990	ArgProArgGlnGlnSerLeuSerLysGluGlySerIleGlyGlySerG1	1006
84590	CGTCCACGGCAACAGAGCCTCAGCAAGGAGGGCAGCATTTGGGGGCAGCGG	84639
1006	yGlySerGlyGlyGlyGlyGlyGlyGlyLeuLysProSerIleThrLys.	1022
84640	GGCAGCGGTGGCGGAGGGGTGGGGGGCTGAAGCCCTCCATTCACCAAGC	84689
1022	.....	1022
84690	AGGTAGGTGAAGGCAGGAGGAAGCGGGCTGGGTCAACAACAGGGAGGGAA	84739
1022	.....	1022
84740	GAAGGACATGGGGGTGGGGTTGAAACAGAGTCTGTGGCCTGAACATTACA	84789
1022	.....	1022





134 nArgAlaValLysProAsnLysAspAsnSerArgArgValAspAsnValL 151  
:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
587 CAGGACAGTTCAACCTAATAAGGACAATTCGAGGGAGCTGAATAATGTC 636

151 euLysLeuTrpIleGluAlaArgGluLeuProLysLysArgTyr 167  
||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
637 TTCGTTTATGGATCATTTGAAGCCAAGGACCTTGCCCTTAAGAAGAAATAT 686

168 TyrCysGluLeuCysLeuAspAspMetLeuTyrAlaArgThrThrSerLy 184  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
687 TTCTGCCAACTGTGCCTTGATGATACCCTCTTTGCTCGTACAACACGCAA 736

184 sProArgSerAlaSerGlyAspThrValPheTrpGlyGluHisPheGluP 201  
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
737 GACCAAAGCA.....GACAATATTTCTGGGGCGAACATTTTGAAT 777

201 heAsnAsnLeuProAlaValArgAlaLeuArgLeuHisLeuTyrArgAsp 217  
||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
778 TCTTCAGCCCTTCCACCTCTTCATAGTATCACAGTTTCACATTACAAGGAT 827

218 SerAspLysLysArgLysLysAspLysAlaGlyTyrValGlyLeuValTh 234  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
828 GTGGAAAAAAGAAAAAAGGACAAGAATAATTATGTAGGGCTAGTCAA 877

234 rValProValAlaThrLeuAlaGlyArgHisPheThrGluGlnTrpTyrP 251  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
878 CATCCCCACTGCCAGTGTGACTGGTCGCCAATTTGTAGAAAAGTGGTATC 927

251 roValThrLeuProThrGlySerGlyGlySerGlyGlyMetGlySerGly 267  
||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
928 CAGTGAGTACACCTACACCCAAC.....AAAAAAAAAGACAGGAGG 967

268 GlyGlyGlyGlySerGlyGlyGlySerGlyLysGlyLysGlyLysGlyCy 284  
.....AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGACAGGAGG 967

284 sProAlaValArgLeuLysAlaArgTyrGlnThrMetSerIleLeuProm 301  
||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
968 ACCTTCTATTTCGGATTAAATCACGTTTCCAAACTATCACCATTTCTGCCTA 1017

301 etGluLeuTyrLysGluPheAlaGluTyrValThrAsnHistyrArgMet 317  
||| ||| ||| ||| ||| ||| :::|:::|:::|:::|:::|:::| |||  
1018 TGGAGCAATACAAAGAAATTGCAGAAATTGTCCACGACCAACTACACCATG 1067

318 LeuCysAlaValLeuGluProAlaLeuAsnValLysGlyLysGluGluVa 334  
||| ||| ||| ||| ||| ||| :::|:::|:::|:::|:::|:::| |||  
1068 CTGTGTTCTGCTCCTTGAGCCAGTAATTAGTGTGAGAAATAAAGAGGAGTT 1117

334 lAlaSerAlaLeuValHisIleLeuGlnSerThrGlyLysAlaLysasp 351  
:||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
1118 GGCTTGTCCTTAGTGCACATTTCTCAAAGTACTGGCAGAGCCCAAGGATT 1167

351 heLeuSerAspMetalamMetSerGluValAspArgPheMetGluArgGlu 367  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
1168 TTCTGACTGACTGGTGATGCTGAGGTGGATCTGTGTGGAGAGCATGAT 1217

368 HisLeuIlePheArgGluAsnThrLeuAlaThrLysAlaIleGluGluTy 384  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
1218 GTCTTGATCTTCAGAGAGAACAATAATGCCACCAATCCATTGAGGAATA 1267

384 rMetArgLeuIleGlyGlnLysTyrLeuLysaspAlaIleGlyGluPheI 401  
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1268 CCTCAAGTTGGTGGACACAACAGTATCTTCATGACGCACCTGGGGGAGTTTA 1317

401 leArgAlaLeuTyrGluSerGluGluAsnCysGluValAspProfileLys 417  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
1318 TCAAAGCTTTGTATGATGCCGATGAGAACTGTGAAGTGGATCCCAGCAA 1367

418 CysThrAlaSerSerLeuAlaGluHisGlnAlaAsnLeuArgMetCysCy 434  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
1368 TGTTTCATCTAGTGAACCTGATAGACCATCACAGCAACCTGAAAATGTGCTG 1417

434 sGluLeuAlaLeuCysLysValValAsnSerHisCysValPheProArgG 451

|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1418 TGAGCTGGCTTTCTGCAAGATCATCAACTCTTACTGTGTTTTCCCTCGTG 1467

luLeuLysGluValPheAlaSerTrpArgLeuArgCysAlaGluArgGly 467  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1468 AGTTGAAGAAGTGTTCATCATGGAAGCAGACAGTGCCTGAACCCTGGC 1517

ArgGluAspIleAlaAspArgLeuIleSerAlaSerLeuPheLeuArgPh 484  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1518 AAGCAAGACATCAGCGAGAGGCTCATCAGTGCCTCATATTCTCCGTTT 1567

eLeuCysProAlaIleMetSerProSerLeuPheGlyLeuMetGlnGluT 501  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1568 TCTGTCTCCAGCCATTATGCTCCCAGTCTTTTCAACCTTATGCAGSAGT 1617

yrProAspGluGlnThrSerArgThrLeuThrLeuIleAlaLysValIle 517  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1618 ATCCTGATGACCGCACATCTCGACTCTAACTCTATTGCCAAGGTCAAT 1667

GlnAsnLeuAlaAsnPheSerLysPheThrSerLysGluAspPheLeuGl 534  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1668 CAGAACCTGSCCAACTTTGCCAAGTTTGGTAACRAAGAGGAATACATGGC 1717

yPheMetAsnGluPheLeuGluLeuGluTrpGlySerMetGlnGlnPheL 551  
:|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1718 ATTTCATGAATGATTTTTAGAACATGAATGGGTGGAATGAAGCGCTTC 1767

euTyrGluIleSerAsnLeuAspThrLeuThrAsnSerSerPheGlu 567  
|| |||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1768 TTTTGGAGATCTCTAATCCAGACACCATCTCAAACACCCCCAGGCTTGAT 1817

GlyTyrIleAspLeuGlyArgGluLeuSerThrLeuHisAlaLeuLeuTr 584  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1818 GGTTCATTGATCTGGCGCGAGAGCTTTCAGTTTTGCATTCCTTACTGTG 1867

pGluValLeuProGlnLeuSerLys.....GluA 594  
|||||::: |||||:::|||  
1868 GGAAGTAGTTCCCAACTTGATAAGGTGAAAATTCCTTCCTACAGCGCA 1917

laLeuLeuLysLeuGlyProLeuProArgLeuLeuSerAspIleSerThr 610  
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1918 CCGTGGCAAAATGGGGCCTTCCTCGTGTCTTGCTGATATTACCAAG 1967

AlaLeuArgAsnProAsn...lleGlnArgGlnProSerArgGlnSerGl 626  
::||| |||||::: |||||:::||| ||| ::|||  
1968 TCATTGACTAATCCTACGCCAATACAACAGCAACTGAGACGCTTCACTGA 2017

uArgAlaArgSerGlnProMetValLeuArgGlyProSerAlaGluMetG 643  
|:::  
2018 ACAT ..... 2021

InGlyTyrMetMetArgAspLeuAsnSerSerIleAspLeuGlnSerPhe 659  
||||||| :::::::::::::::  
2022 .....AACTCCAGTCCAATGTCAAGTGAAGC 2048

MetAlaArgGlyLeuAsn.....SerSerMetAspme 670  
::: |||||:::  
2049 CTCTCCTCTGGGCTGCAGAAAAATATTGAAGACCCCACTGACAGTGATTT 2098

tAlaArgLeuProSerProThrLysGluLysProProProProProg 687  
: :||| | |||:::|||||:::|||||:::|||||  
2099 GCATAAACAATAAAATCTCCAAGCCAGGACAAC.....ACAGACAGCTACT 2142

lyGlyGlyLysAspLeuPheTyrValSerArgProProLeuAlaArgSer 703  
||||| ||::: ||:::|||||:::|||||  
2143 TCAGAGGGGAAACATATTATGCTGGTTCAGCAA.....GCCTCC 2180

SerProAlaTyrCysThrSerSerSerAspIleThrGluProGluGlnLy 720  
||| ::: |||:::|||||  
2181 TCTCAGACATGACTTATTCTGAAAAGGAT.....GAAAGGGAAG 2221

sMetLeuSerValAsnLysSerValSerMetLeuAspLeuGlnGlyAspG 737  
: ||| :::::|||||:::|||||:::|||||:::|||||

2222 TAGCCTTCCTAATGGTCGGAGCGCTCCCTCATGGACCTCCAG...GACA 2268  
737 lyProGlyGlyArgLeuAsnSerSerSerValSerAsnLeuAlaVal 753  
2269 CTCATGCTGCTCAAGTGGAGCATGCATCTGTC.....ATGCTTGATGTG 2312  
754 GlyAspLeuLeuHisSerSerGlnAlaSerLeuThrAlaAlaLeuGlyLe 770  
2313 CCTATACGCTTGACCGGAAGCCAGCTTCCATAACCCAGGTGGCCAGCAT 2362  
770 uArgProAlaProAlaGlyArgLeuSerGlnGlySerGlySerSerIlet 787  
2363 CAAA.....CAGCTGGGGAACCCAGAGCACTCCCCAAA 2397  
787 hrAlaAlaGlyMetArg.....LeuSerGlnMetGly 797  
2398 GTGCACCCCAAGTGAGAGGCCCTCGCACCCAGCCTTGAACCCAGCCAGGT 2447  
798 ValThrThrAspGlyValProAlaGlnGlnLeuArgIleProLeuSerPh 814  
2448 .....GGCCTTCAG.....CCCTTGCTCGTT 2467  
814 eGlnAsnProLeuPheHisMetAlaAlaAspGlyProGlyProProAlaG 831  
2468 CCAGAACCTGTCTATCACCTCAATAACCCCAATTCCAGCAATGCCAAAG 2517  
831 lyHisGlyGlySerSerGlyHisGlyProProSerSerHisHisHis 847  
2518 CCTCTATAGATTCCAGT..... 2534  
848 HisHisHisHisHisArgGlyGlyGluProProGlyAspThrPheAl 864  
2534 ..... 2534  
864 aProPheHisGlyTyrSerLysSerGluAspLeuSerThrGlyValProL 881  
2535 .....TTGGAGAACCCTAAGCACT..... 2552  
881 ysProProAlaAlaSerIleLeuHisSerHisSerTyrSerAspGluPhe 897  
2553 .....GCCAGTTCC.....AGAGCCAAAGTAACAGTGAAGACTTC 2588  
898 .....GlyProSerGlyThr.....AspPheThrArgArgG1 908  
2589 AAGCTCAGTGGACCCAGCAATAGCAGCATGGAAGATTTCCTAAACGTAG 2638  
908 nLeuSerLeuGlnAsp...AsnLeuGlnHisMetLeuSerProProGlnI 924  
2639 CACTCAGAGTGAGGACTTCTCCAGCGGCACACGGTGTCCAGATAGACACA 2688  
924 leThrIleGlyProGlnArgProAlaProSerGlyProGlyGlySer 940  
2689 TACCTCTGTGTTTGCACGACAAAATAGTACTGGGCGAGGCCAGATCCGA 2738  
941 GlyGlyGlySerGlyGlyGlyGly.....GlyGlyGlnProProProLe 955  
2739 AAAGTGGACCAAGGTGGGTAGGTGCCCCGAGCCAAAGCCCCCACCATCCCT 2788  
955 uGlnArgGlyLysSerGlnGlnLeuThr.....ValSerA 967  
2789 GCCACACAGTGCTTCTTACGTAGCACCCGGGAGCATGTGAGTGGTCCG 2838  
967 laAlaGlnLysProArgProSerSerGlyAsnLeuLeuGlnSerProGlu 983  
2839 CAGCC.....CTGGTGGCCGAA 2855  
984 ProSerTyrGlyProAlaArgProArgGlnGln...SerLeuSerLysG1 999  
2856 CCTGTGCAGAATGGGAGCGGTCCCGGCAGAGTCTCTCTCTCCAGAGA 2905  
999 uGlySerIleGlyGlySerGlyGlySerGlyGlyGlyGlyGlyGlyL 1016  
2906 GAGCCCTGTT..... 2915

1016 euLysProSerIleThrLys...GlnHisSerGlnThrProSerThrLeu 1031  
2916 .....CCCAAAGTTAGAGCAATCCAGAGACAACAGACAGCAGGTTTCAG 2960  
1032 AsnPro.....ThrMetProAlaSerGluArgThrValAlaTr 1044  
2961 TCACCTGTGGACTCTGCCACATGTCCCCAGTAGAGAGGACAGCAGCCTG 3010  
1044 pValSerAsnMetProHisLeuSerAlaAspIleGluSerAlaHisIleG 1061  
3011 GGTCTGAACAATGGCAGTATGAAGAGGATGTGGAA..... 3047  
1061 luArgGluGluTyrLysLeuLysGluTyrSerLysSerMetAspGluSer 1077  
3048 .....GAAACTGAGCAAAATCTAGATGAAGCC 3074  
1078 ArgLeuAspArgValLysGluTyrGluGluGluIleHisSerLeuLysG1 1094  
3075 AAG.....CATGCTGAGAAGTATGAACAAGAAATTACTAACTGAAGGA 3118  
1094 uArgLeuHisMetSerAsnArgLysLeuGluTyrGluArgArgLeuL 1111  
3119 GCGCCTGAGAGTTTCCAGCCCGGCGACTGGAGGAATATGAACGCCGCTTC 3168  
1111 euSerGlnGluGlnThrSerLysIleLeuMetGlnTyrGlnAlaArg 1127  
3169 TGGTGCAGGAGCAGCAGATGCAGAAAGCTGCTGCTGGAATACAAGGCCGA 3218  
1128 LeuGluGlnSerGluLysArgLeuArgGlnGlnValGluLysAspSe 1144  
3219 CTGGAGGACAGCGAGGAGCGGCTCCGAAGACAGCAGGAGGAGAAAGATAG 3268  
1144 rGlnIleLysSerIleGlyArgLeuMetLeuValGluGluLeuA 1161  
3269 CCAGATGAAAGCATCATCAGCAGCGCTAATGGCTGTGGAAGAGGAACTGA 3318  
1161 rgArgAspHisProAlaMetAlaGluProLeuProGluProLysLysArg 1177  
3319 AGAAGGATCATGCTGAGATG...CAAGCAGTTATTGATGCAAAAGCAGAAA 3365  
1178 LeuLeuAspAlaGln.ArgGlySerPheProProTyrValGlnGlnThra 1194  
3366 ATAATTGATGCACAGGAAACACGGATCGTGTCCCTGGATTCAGCCAACAC 3415  
1194 rg.....ValArgTrpPro.....HisLeuGlyThrAlaTrp... 1204  
3416 CAGACTGATGAGCGCGCTGACCCCAAGTGAAGGAGCGGTACAGCATGCAGG 3465  
1205 .....ProProGlnProHisProProHisProGlyCysar 1216  
3466 TCCGCAATGGCATCTCCCCCACCACCCCAAGCTTCCATCAGCGAG 3515  
1216 g.....SerGlnArgThrAlaSerSerGlyThr..... 1225  
3516 AATGGTGAATTCAAAACACAGCAGCTGCTGACGGGCTTGTGTGTGGAAGG 3565  
1226 .....ProGlnThrThrSerProProSerIleThrAsp 1236  
3566 AGACAGAAGGAATTGACCCCACTCTCTATCTCCA.....GAC 3603  
1237 LeuLeuProCysAlaProTyrProGlyProProSerValThrAspLeuLe 1253  
3604 CTT.....TACCTAGCCCCCTCCAGGTTTACAGAAATGTTGC 3638  
1253 uProSer 1255  
3639 TACTTCA 3645

seq\_name: gb\_htg7:AC019800

seq\_documentation\_block:

LOCUS

DNA

56366 bp

HTG

03-JAN-2000

DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.

ACCESSION AC019800  
VERSION AC019800.1 GI:6665097  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 56366)

AUTHORS Adams,M. and Venter,J.C.

TITLE Direct Submission

JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,

Rockville, MD, USA

COMMENT This sequence was identified as CDM:10210991 by the submitter.

For more information on this record e-mail to fly@celera.com.

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

FEATURES Location/Qualifiers

source  
1..56366  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"

BASE COUNT 16234 a 11924 c 11808 g 16400 t

ORIGIN

alignment\_scores:

Quality: 1257.50 Length: 1432  
Ratio: 1.642 Gaps: 57  
Percent Similarity: 53.492 Percent Identity: 29.190

alignment\_block:

US-09-294-298-6 x AC019800/rev ..

Align seg 1/1 to reverse of: AC019800 from: 1 to: 56366

19 PheLeuSerArgArgLeuLysSerSerIleLysArgThrLysSerGlnPr 35

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

26122 TTCCTTTGCCACAGA...TCCAATCCTTTGAAGCGGACCAAGTCGTGAC 26076

35 oLysLeuAspArgThrSerSerPheArgGlnIleLeuProArgPheArgS 52

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

26075 CAAGCTGGAGCGGACCAAG..... 26057

52 erAlaAspHisAspArgAlaArgLeuMetGlnSerPheLysGluSerHis 68

||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

26056 .....CGCGGATCCGGCGGACTGAGGGGCTCCCGC 26027

69 SerHisGluSerLeuLeuSerProSerSerAlaAlaGluAla..... 82

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

26026 TCGCAGGAGAGTCTGCTGTCAGTCACGCCGTCATGTCACCATAGGTGA 25977

82 ..... 82

25976 GTTCTGGCTCATAATGGAGTTTGGCCAAACCCCTCAAACCATTTCTTT 25927

83 .....LeuGluLeuAsnLeuAspGluAspSerIleLysPro 95

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25926 TCTTCTCTTCTAGATCTCTCGTGCACCTGGGCGGTGGCGTGGCGCCC 25877

96 ValHisSerSerIleLeuGlyGlnGluPheCysPheGluValThrSe 112

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25876 GTGCATCAGTCGTTCTGGGACGTCGTCTACTGTTCCAGGTACGGGCGG 25827

112 r...SerGlyThrLysCysPheAlaCysArgSerAlaAlaGluArgAspL 128

||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25826 GCCTCGTGGCGAGCGGTACTCATCGGATCGCGCCAGGAGCGGACCC 25777

128 yStrpIleGluAsnLeuGlnArgAlaValLysProAsnLysAspAsnSer 144

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25776 TTTGGATCTACTCGCTGCGCAAGTCGATCGCTCCGAATGCAGACACAG 25727

145 ArgArgValAspAsnValLeuLysLeuTrpIleIleGluAlaArgGluLe 161

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25726 CGTCGCACGACAACCTCGCTGAGATGTGGGTGTACGAGCGGAAAAATCT 25677

161 uProProLysLysArgTyrTyrCysGluLeuCysLeuAspMetLeuT 178

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25676 GCCGCCCAAGAAGCGTTACTTTTGGAACTGCAATTGGACAAGACGCTGT 25627

178 yAlaArgThrThrSerLysProArgSerAlaSerGlyAspThrValPhe 194

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25626 ACGGCCGACTTCGGTGAAGCTGCAGACG.....GATCTGCTGTTT 25586

195 TrpGlyGluHisPheGluPheAsnAsnLeuProAlaValArgAlaLeuAr 211

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25585 TGGGGGAGCAGCTTCGATTTCGCCGACATACCCGAGATTAAATGTGATCAC 25536

211 gLeuHisLeuTyrArgAspSerAspLysLysArgLysLysAspLysAlaG 228

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25535 TGTTAACGTTTTCGGTGAGGTGGACAAGAAGACGGGGACAAATACC 25486

228 lyTyrValGlyLeuValThrValProValAlaThrLeuAlaGlyArgHis 244

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25485 AATTGTGGGATCGGTGAAGATACCCGTGCACGATGTACACCTCCAGATTG 25436

245 PheThrGluGlnTrpTyrProValThrLeuProThrGlySerGlyGlySe 261

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25435 CCCTGCGAACAATGGTATCCGATA...CTGAGCGACAAAGCGGGCGACAG 25389

261 rGlyGlyMetGlySerGlyGlyGlyGlySerGlyGlyGlySerGlyG 278

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25388 TCTGGGCGAGGACATCGGGC.....GCGGGCGGCGAGTGGGT 25354

278 lyLysGlyLysGlyGlyCysProAlaValArgLeuLysAlaArgTyrGln 294

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25353 CCAAGGACAGGAGCAATTGCCACGCTGAGGATCAAGTGTCTGTTCCAG 25304

295 ThrMetSerIleLeuProMetGluLeuTyrLysGluPheAlaGluTyrVa 311

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25303 AGCACCGAGATCCTGCCATCAATGTGTACGGCACTTTTTCAGCTACCT 25254

311 lThrAsnHisTyrArgMetLeuCysAlaValLeuGluProAlaLeuAsnV 328

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25253 CAAGGAGAACTATAAGCGCGTGTGCGAGACCCCTGGAGCCCGGTGATCGGAG 25204

328 alLysGlyLysGluGluValAlaSerAlaLeuValHisIleLeuGlnSer 344

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25203 TCAAGGCCAAGGAGGACATTGGACAGGCACCTGCTGCTGATGCACGCA 25154

345 ThrGlyLysAlaLysAspPheLeuSerAspMetAlaMetSerGluValas 361

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25153 CAGGGATTGGCGGGCGCCCTTCCTCACCAGTGTGGTGGCCCTCGATCTGCT 25104

361 pArgPheMetGluArgGluHisLeuIlePheArgGluAsnThrLeuAlaT 378

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25103 GCGAGTTGGCGAT...CAGAGCTTACGTTACGGGGCAACTCCTTGGCCA 25057

378 hrLysAlaIleGluGluTyrMetArgLeuIleGlyGlnLysTyrLys 394

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25056 CCAAGAGCATGGAGGCAATTCCTCAAGCTGACGGCGCAACAGTATCTGCAG 25007

395 AspAlaIleGlyGluPheIleArgAlaLeuTyrGluSerGluGluAsnCy 411

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

25006 GACACACTATCCGCCACCAATAAACGAGCTAATTACGTCGGAGAGGACTG 24957

411 sGluValAspProIleLys.....CysThrAlaSerSerLeuAlaG 425

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

24956 CGAGGTGGATCCCAACCAAGACGAGCGGTTCGTCGCGGGTTCGCTGCAGC 24907

425 luHisGlnAlaAsnLeuArgMetCysCysGluLeuAlaLeuCysLysVal 441

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

24906 GACAGCAGCGCCCTTGGCTGGCGGGTCCGAGGGGCGGTGGCAGTGCATC 24857



442 ValAsnSerHisCysValPheProArgGluLeuLysGluValPheAlaSe 458  
 24856 TTCGAATCGCACAAAGCATTTCCCGCCCGAGTTACGAAATTCGTTGCGAC 24807  
 458 rTrpArgLeuArgCysAlaGluArgGlyArgGluAspIleAlaAspArgL 475  
 24806 GTTCCGGGAGCGCTTGCACAGCTGGCCCGTCAGGATATGGCCGACAACC 24757  
 475 euIleSerAlaSerLeuPheLeuArgPheLeuCysProAlaIleMetSer 491  
 24756 TGATCTGGCGGAGCATTTCTCGGCTTCTGTCGTCGCGGCGCATCTGTGCG 24707  
 492 ProSerLeuPheGlyLeuMetGlnGluTyrProAspGluGlnThrSerAr 508  
 24706 CCGTCGCTCTTCAATATCACCAGCGAACTGCGCTCCGCGCACGGGCCACCCG 24657  
 508 gThrLeuThrLeuIleAlaLysValIleGlnAsnLeuAlaAsnPheSerL 525  
 24656 CAATCTCACACTGGTGGCCAAAGACCCTGCAACACATTGGCCCAATTTCACCC 24607  
 525 ySPheThrSerLysGluAspPheLeuGlyPheMetAsnGluPheLeuGlu 541  
 24606 GCTTCCAGGGCAAGAGAACTTTATGGAGTTTCTCAACGATTTCCTCGAG 24557  
 542 LeuGluTrpGlySerMetGlnGlnPheLeuTyrGluIleSerAsnLeu.. 557  
 24556 CAGGAGCGCGCTCGCATGCAACAGTTTCTGGAGATTATATCCACCGGCC 24507  
 558 .....AspThrLeuThrAsnSerSerPheGluGlyT 569  
 24506 GGAGCACCCAGCTCCAGACTCGATCTCGAT.....TGGCGCGGGT 24466  
 569 yrIleAspLeuGlyArgGluLeuSerThrLeuHisAlaLeuLeuTrpGlu 585  
 24465 ACATCGACAGGGCAACACAGTTGTCCATACACAGTTTGTCTCAGCGAA 24416  
 586 valLeuProGlnLeuSerLysGluAlaLeuLeuLysLeuGlyProLeuPr 602  
 24415 AGCCTGGCCAAAGCTGCCGGAGGCCAGGCAGCAGCTGGATCCGTTGCA 24366  
 602 oArgLeuLeuSerAspIleSer.....ThrA 611  
 24365 GCATATTCTCGATGAATCAGCCGAGCCAAAGAGCATGGCATGGGCACAG 24316  
 611 laLeu..... 612  
 24315 CACTGCCGGGTGGATATTTGCGGGCCACCTCGTCTACGCACCTCGATAGCC 24266  
 613 .....ArgAsnPro..... 615  
 24265 AGCGAGAATCAGGAGAATCGCAATCCGGGATCTCTGGGCTCGCACCGCTGG 24216  
 616 ....AsnIleGlnArgGlnProSerArgGlnSerGluArgAlaArgSerG 631  
 24215 CTCCAACCTCGGAGCAGTTACTGCCACAAAGCCAGTTGGCCCGCCGCGC 24166  
 631 lnProMetValLeuArgGlyPro...SerAlaGlu.....MetGln 643  
 24165 AGCATGCGATGTTAGTAACCATTTATCTGCGGAGCGCGGCATCATCGGA 24116  
 644 Gly.TyrMetMetArgAspLeuAsnSerSerIleAspLeuGlnSerPheM 660  
 24115 GGAGTACTACGCCGAATTCTCTGGAGAAGATAATCTTTAGATACATGA 24066  
 660 etAlaArgGlyLeuAsnSerSerMetAspMetAlaArgLeuProSerPro 676  
 24065 TCCACCGGTTAATG..... 24052  
 677 ThrLysGluLysProProProProGlyGlyLysAspLeuPh 693  
 24052 ..... 24052

693 eTyrValSerArgProProLeuAlaArgSerSerProAlaTyrCysThrS 710  
 24051 .....GCTTACTGCAGCA 24039  
 710 erSerSerAspIleThrGluProGluGlnLysMetLeuSerValAsnLys 726  
 24038 GCAGCAGCAGCAGCAGCAGC.....AGCAGCAACAAC 24007  
 727 SerValSerMetLeuAspLeuGlnGlyAspGlyProGlyGlyArgLeuAs 743  
 24006 AGCAACAGCAACATC.....AGCAGCTGCAA 23981  
 743 nSerSerSerValSerAsn.....LeuAlaAlaValGly..AspLeuLe 757  
 23980 CAGCATGGCCATCAGCAACAGCCGCGCACCCAGCATCCACTCCAGATGCT 23931  
 757 uHisSerSerGlnAlaSerLeuThrAlaAlaLeuGlyLeuArgProAlap 774  
 23930 CTCCAATTCAAAACCTCCATT...GCCGGCAACCAATATATGAGTTCCG 23884  
 774 roAlaGlyArgLeuSerGlnGlySerGlySerSerIleThrAlaAlaGly 790  
 23883 CAGGAGCGCTGCAGCATGCCCAATCGCAGACCTCGATGGCGTCTCATCG 23834  
 791 MetArgLeuSerGlnMetGlyValThrThrAspGlyValProAlaGlnGl 807  
 23833 CTT..... 23831  
 807 nLeuArgIleProLeuSerPheGlnAsnProLeuPheHisMetAlaAlaA 824  
 23831 ..... 23831  
 824 spGlyProGlyProProAlaGlyHisGlyGlySerSerGly.....His 838  
 23830 .....AATGGGAGCAGCAGCAATTGCTGCAC 23804  
 839 GlyProProSerSerHisHisHisHisHisHis.....Hi 851  
 23803 GGC.....CACCAGCAGCATGCCCATCACCCGCGCAACTGCA 23766  
 851 sHisHisArgGlyGlyGluProProGlyAspThrPheAlaProPheHisG 868  
 23765 TCCACATCACTGC.....CCGCCG..... 23747  
 868 lyTyrSerLysSerGluAspLeuSerThrGlyValProLysProProAla 884  
 23746 .....GCCACAGCAGCAGTGCC 23729  
 885 AlaSerIleLeuHis.....SerHisSerTyrSerAspGl 896  
 23728 TCCAGCACTATGGAGCGCATGGATCGCATGAACATATCCGTATATGTCGCA 23679  
 896 uPheGlyProSerGlyThrAspPheThrArgArgGlnLeuSerLeuGlnA 913  
 23678 C.....AATGGCAATGACTACGAG..... 23660  
 913 spAsnLeuGlnHisMetLeuSerProProGlnIleThrIleGlyProGln 929  
 23659 .....ACCAGCAGCGCTTCGAGCAGCTCGCTCCAGGACA 23627  
 930 ArgProAlaProSerGlyPro.....GlyGlyGlySerGlyGlySe 944  
 23626 CTGCCACGGAATGGAATCCCAATGCCAATGGCAACGTGGGCGCAGCAAA 23577  
 944 rGlyGlyGlyGlyGlnProProProProProProProProProProPro 960  
 23576 CAATAACAGAGCGCGCAGTACGATGACATGACGCGGAGTTCCAAATCC 23527  
 961 GlnGlnLeu.....ThrValSerAlaAlaGlnLysProAr 972  
 23526 AGATCTCTGGGTTCGATACGAGCAGTGTCTTGTCTGCAAGTCGCCACACA 23477  
 972 gPro.....Ser.SerGlyAsnLeuLeuGlnSerPro 982

23476 CCCATGATGAATCCAGTTTGGACACGCGGGCCGCGGAGTCAATCA 23427  
::: |||||::: ::::|  
983 GluProSerTyrGlyProAlaArgProArg..... 992  
::: ::::|::: ::::|  
23426 CAAACTGAATTTGGGAATACCCGATCACTCAGTGCGCTATGTGCGGGTA 23377  
::: ::::|::: ::::|  
993 .....GlnGlnSerLeuSerLysGluGlySerIleGlyGlySerGlyG 1007  
::: ::::|::: ::::|  
23376 ATAATTTGAATCCCAACTCGAATATGCCCAAGAACTTGGAGGATCTGGAC 23327  
::: ::::|::: ::::|  
1007 lySer.....GlyGlyGlyGlyGlyGlyGlyGlyLeuLysPro 1018  
||| |||||  
23326 GATCTGTTCAAGTACGCCGAGGAGCATGACGTGCGCGGAACCAACCA 23277  
||| |||||  
1019 SerIleThrLysGlnHisser...GlnThrProSerThr. LeuAsnPro. 1033  
||| ::::| ||| ::::| |||||  
23276 TCACAATCACACCAAGGTCAGCAGAACCAACAGGACATCTGAAGCCGG 23227  
||| |||||  
1034 ..ThrMetProAlaSerGluArgThrValAlaTrpValSerAsnMetPro 1049  
::: ::::| |||||  
23226 CCGCCGTTCCCGCAAGGAGCAGCTGTCGCGGAAAGCAGT..... 23186  
||| |||||  
1050 HisLeuSerAlaAspIleGluSerAlaHisIleGluArgGluGluTyrLy 1066  
||| |||||  
23185 CACTGCAGTTCTGGCTACACAGAGCATCTCCACAAATCCC..... 23147  
||| |||||  
1066 sLeuLysGluTyrSerLysSerMetAspGluSerArgLeuAspArgVal 1083  
||| |||||  
23146 .....TCGCCCTCGCAGTCTCCAGTCCCGTGGAG..... 23117  
||| |||||  
1083 ysGluTyrGluGluGluIleHisSerLeuLysGluArgLeuHisMetSer 1099  
::: |||||  
23116 .....AGCCAGCTGAAGCGCGCGATGGCAGTCAC 23087  
||| |||||  
1100 AsnArgLysLeuGluGluTyrGluArgArgLeuLeuSerGlnGluGlu 1116  
||| |||||  
23086 AATGCGCGCGTG...GCCTTCAAGAATCCCTCCTATCAGCTTCAGCCCA 23040  
||| |||||  
1116 nThrSerLysIleLeuMetGlnTyrGlnAlaArgLeuGluGlnSerGlu 1133  
||| |||||  
23039 AACTGGCTCGTCCAGATCATCGGCACAGAGTAATACACACAGCAGCAGC 22990  
||| |||||  
1133 ysArgLeuArgGlnGlnValGluLysAspSerGlnIleLysSerIle 1149  
::: |||||  
22989 AGCAGCAGCAACAACACAGTTT.....GGCAGCGCTTGAACCAATT 22946  
||| |||||  
1150 IleGlyArgLeuMetLeuValGluGluGluLeuArgArgAspHisProAl 1166  
||| |||||  
22945 GCGGCGGACTGTGGCGCGGAGCGGCTTCTCTCAACAGTGGCGGAGC 22896  
||| |||||  
1166 aMet.AlaGluProLeu.....ProGluProLys..... 1175  
||| |||||  
22895 CTTGGAGCGCGCCACTTTGACGCCAGCTCTCTCGGACGAACAGCTCTCGG 22846  
||| |||||  
1176 .....LysArgLeuLeuAspAlaGlnArgGlySe 1185  
::: |||||  
22845 CGGATAATTACTTCAATTATGACGCGGCTGCAGCTGCTGGAGCGGTATT 22796  
||| |||||  
1185 rPheProProTrpValGlnGlnThrArgValArgTrpProHisLeuGlyT 1202  
||| |||||  
22795 GCCACCAATTGGAGGCTCAACGCTCGC.....TCAGCGCGGCGCAG 22755  
||| |||||  
1202 hrAlaTrpProProGln..... 1207  
::: |||||  
22754 TAGCTCCTCCACCTCAGCATCTCGCTCCACCTCGAATCTTGGCAAGAGCG 22705  
||| |||||  
1208 ...ProHisProProHisProGlyCysArgSerGlnArgThrAlaSerSe 1223  
||| |||||  
22704 GCGGCTCATCCGCCTACGGCGGCTGAATGGCGCGCTGAAGCGCGGAGGAT 22655  
||| |||||  
1223 rGlyThrProGlnThrThrSerProPro.....SerIleThrA 1236  
::: |||||

22654 GTCTACGGCAGTGGCTACGGCGGACGAGTGGAAATGTGGCTATGGCTT 22605  
1236 spLeuLeuPro.Cys.....AlaProTyrProGlyProProSerValTh 1250  
||||| ||| ::::|  
22604 GTCCACTTCCAGTGGCGGCGGACACCATCAACATCCCCACGACGAGC 22555  
||| |||||  
1250 rAspLeuLeuProSerAlaProAspProGlyThrSerProThrThrArgT 1267  
::: ::::| |||||  
22554 AGAATCCGATGCGACGACGACGACAAAGGGAACGGGATCAGGAACACAAG 22505  
||| |||||  
1267 hr.....GlyArgHisGlnGlyThrAlaGlyLeuSerPro 1278  
::: |||||  
22504 CAGTATGCCGCGAGTGTGGCGGCGGACGCTGGGATCGGCCACATCAGCGGC 22455  
||| |||||  
1279 LeuThrProPro.....TrpGlyThrHis 1286  
||| |||||  
22454 TCAGAGGCGCTGAGCTTGGACTCGGCGCGGCAC 22422  
||| |||||  
seq\_name: gb\_htg4:AC012162  
seq\_documentation\_block:  
LOCUS AC012162 191504 bp DNA HTG 09-FEB-2000  
DEFINITION Drosophila melanogaster chromosome X clone BACR01N10 (D1115)  
RPCI-98 01.N.10 map 16D-16F strain Y; cn bw sp, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 142 unordered pieces.  
ACCESSION AC012162  
VERSION AC012162.9 GI:6957580  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 191504)  
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,  
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,  
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,  
Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,  
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.  
Sequencing of Drosophila melanogaster  
Unpublished  
2 (bases 1 to 191504)  
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,  
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,  
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.  
Direct Submission  
Submitted (21-OCT-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Feb 11, 2000 this sequence version replaced gi:6838825.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdgpefruitfly.berkeley.edu](mailto:bdgpefruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 142 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 456: contig of 456 bp in length  
\* 457 536: gap of unknown length

\* 537 1004: contig of 468 bp in length  
\* 1005 1084: gap of unknown length  
\* 1085 1958: contig of 874 bp in length  
\* 1959 2038: gap of unknown length  
\* 2039 2248: contig of 210 bp in length  
\* 2249 2328: gap of unknown length  
\* 2329 2858: contig of 530 bp in length  
\* 2859 2938: gap of unknown length  
\* 2939 3325: contig of 387 bp in length  
\* 3326 3405: gap of unknown length  
\* 3406 4165: contig of 760 bp in length  
\* 4166 4245: gap of unknown length  
\* 4246 4716: contig of 471 bp in length  
\* 4717 4796: gap of unknown length  
\* 4797 5373: contig of 577 bp in length  
\* 5374 5453: gap of unknown length  
\* 5454 6050: contig of 597 bp in length  
\* 6051 6130: gap of unknown length  
\* 6131 6720: contig of 590 bp in length  
\* 6721 6800: gap of unknown length  
\* 6801 7461: contig of 661 bp in length  
\* 7462 7541: gap of unknown length  
\* 7542 8104: contig of 563 bp in length  
\* 8105 8184: gap of unknown length  
\* 8185 8505: contig of 321 bp in length  
\* 8506 8585: gap of unknown length  
\* 8586 9227: contig of 642 bp in length  
\* 9228 9307: gap of unknown length  
\* 9308 10091: contig of 784 bp in length  
\* 10092 10171: gap of unknown length  
\* 10172 11099: contig of 928 bp in length  
\* 11099 11179: gap of unknown length  
\* 11180 11747: contig of 568 bp in length  
\* 11748 11827: gap of unknown length  
\* 11828 12767: contig of 940 bp in length  
\* 12768 12847: gap of unknown length  
\* 12848 13398: contig of 551 bp in length  
\* 13399 13478: gap of unknown length  
\* 13479 14007: contig of 529 bp in length  
\* 14008 14087: gap of unknown length  
\* 14088 14867: contig of 780 bp in length  
\* 14868 14947: gap of unknown length  
\* 14948 15562: contig of 615 bp in length  
\* 15563 15642: gap of unknown length  
\* 15643 16859: contig of 1217 bp in length  
\* 16860 16939: gap of unknown length  
\* 16940 17986: contig of 1047 bp in length  
\* 17987 18066: gap of unknown length  
\* 18067 19341: contig of 1275 bp in length  
\* 19342 19421: gap of unknown length  
\* 19422 20926: contig of 1505 bp in length  
\* 20927 21006: gap of unknown length  
\* 21007 22153: contig of 1147 bp in length  
\* 22154 22233: gap of unknown length  
\* 22234 23236: contig of 1003 bp in length  
\* 23237 23316: gap of unknown length  
\* 23317 24607: contig of 1291 bp in length  
\* 24608 24687: gap of unknown length  
\* 24688 25620: contig of 933 bp in length  
\* 25621 25700: gap of unknown length  
\* 25701 26300: contig of 600 bp in length  
\* 26301 26380: gap of unknown length  
\* 26381 27488: contig of 1108 bp in length  
\* 27489 28215: contig of 647 bp in length  
\* 27569 28295: gap of unknown length  
\* 28216 29440: contig of 1145 bp in length  
\* 28296 29441 30587: gap of unknown length  
\* 29441 30587: contig of 1067 bp in length  
\* 30588 30667: gap of unknown length  
\* 30668 32042: contig of 1375 bp in length  
\* 32043 32122: gap of unknown length  
\* 32123 33740: contig of 1618 bp in length

33741 33820: gap of unknown length  
33821 35375: contig of 1555 bp in length  
35376 35455: gap of unknown length  
35456 36258: contig of 803 bp in length  
36259 36338: gap of unknown length  
36339 38003: contig of 1665 bp in length  
38004 38083: gap of unknown length  
38084 39509: contig of 1426 bp in length  
39510 39589: gap of unknown length  
39590 40719: contig of 1130 bp in length  
40720 40799: gap of unknown length  
40800 42453: contig of 1654 bp in length  
42454 42533: gap of unknown length  
42534 43841: contig of 1308 bp in length  
43842 43921: gap of unknown length  
43922 45410: contig of 1489 bp in length  
45411 45490: gap of unknown length  
45491 46782: contig of 1292 bp in length  
46783 46862: gap of unknown length  
46863 48570: contig of 1708 bp in length  
48571 48650: gap of unknown length  
48651 51042: contig of 2392 bp in length  
51043 51122: gap of unknown length  
51123 52760: contig of 1638 bp in length  
52761 52840: gap of unknown length  
52841 54712: contig of 1872 bp in length  
54713 54792: gap of unknown length  
54793 56521: contig of 1729 bp in length  
56522 56601: gap of unknown length  
56602 58938: contig of 2337 bp in length  
58939 59018: gap of unknown length  
59019 60837: contig of 1819 bp in length  
60838 60917: gap of unknown length  
60918 62988: contig of 2071 bp in length  
62989 63068: gap of unknown length  
63069 65583: contig of 2515 bp in length  
65584 65663: gap of unknown length  
65664 68774: contig of 3111 bp in length  
68775 68854: gap of unknown length  
68855 71235: contig of 2381 bp in length  
71236 71315: gap of unknown length  
71316 74113: contig of 2798 bp in length  
74114 74193: gap of unknown length  
74194 77405: contig of 3212 bp in length  
77406 77485: gap of unknown length  
77486 80936: contig of 3451 bp in length  
80937 81016: gap of unknown length  
81017 83357: contig of 2341 bp in length  
83358 83437: gap of unknown length  
83438 87456: contig of 4019 bp in length  
87457 87536: gap of unknown length  
87537 91532: contig of 3996 bp in length  
91533 91612: gap of unknown length  
91613 97143: contig of 5531 bp in length  
97144 97223: gap of unknown length  
97224 105593: contig of 8370 bp in length  
105594 105673: gap of unknown length  
105674 112071: contig of 6398 bp in length  
112072 112151: gap of unknown length  
112152 121011: contig of 8860 bp in length  
121012 121091: gap of unknown length  
121092 143113: contig of 22022 bp in length  
143114 143193: gap of unknown length  
143194 143587: contig of 394 bp in length  
143588 143667: gap of unknown length  
143668 143927: contig of 260 bp in length  
143928 144007: gap of unknown length  
144008 144842: contig of 835 bp in length  
144843 144922: gap of unknown length  
144923 145360: contig of 438 bp in length  
145361 145440: gap of unknown length  
145441 145915: contig of 475 bp in length  
145916 145995: gap of unknown length



```

* 145996 146532: contig of 537 bp in length
* 146533 146612: gap of unknown length

alignment_scores:
  Quality: 1257.50      Length: 1432
  Ratio: 1.642          Gaps: 57
  Percent Similarity: 53.492  Percent Identity: 29.190

alignment_block:
US-09-294-298-6 x AC012162 ..

Align seg 1/1 to: AC012162 from: 1 to: 191504

19 PheLeuSerArgArgLeuLysSerSerIleLysArgThrLysSerGlnPr 35
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
113043 TTCTTTGCCACAGA...TCCAATCCTTTGAAGCGGACCAAGTCGGTGAC 113089

35 oLysLeuAspArgThrSerSerPheArgGlnIleLeuProArgPheArgS 52
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
113090 CAAGCTGGAGCGGACCAAG..... 113108

52 erAlaAspHisAspArgAlaArgLeuMetGlnSerPheLysGluSerHis 68
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
113109 .....CGCGGATCCGGCGGACTGAGGGGCTCCCGC 113138

69 SerHisGluSerLeuLeuSerProSerSerAlaAlaGluAla..... 82
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
113139 TCGCACGAGAGTCTGCTGTCCAGTACGCGCGTCAATGCCACCATAGGTGA 113188

82 ..... 82

113189 GTTCTGGCTCATAATGGAGTTTGGGCCAAACCTCAAACCATTTTCTTTT 113238

83 .....LeuGluLeuAsnLeuAspGluAspSerIleIleLysPro 95
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
113239 TCTTCTCTCTTAGATCTCTCGTGCACCTGGGGCGGTGGCGGCC 113288

96 ValHisSerSerIleLeuGlnGluPheCysPheGluValThrThrSe 112
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
113289 GTGCATCAGTCGGTTCTGGGACGTCGTCACTGTTTCCAGGTACGGGGCGG 113338

112 r...SerGlyThrLysCysPheAlaCysArgSerAlaAlaGluArgAspL 128
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
113339 GCCTCGTGGCGAGCGGTACTACTCATCGGATCGCGCCAGGAGCGCGACC 113388

128 ystrPileGluAsnLeuGlnArgAlaValLysProAsnLysAspAsnSer 144
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
113389 TTTGGATCTACTCGCTCGGCAAGTCGATCGCTCCGAATGCAGAGCACACG 113438

145 ArgArgValAspAsnValLeuLysLeuTrpIleIleGluAlaArgGluLe 161
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
113439 CGTCGCACGGACAACCTCGCTGAAGATGTGGGTGTACGAGGCGGAAAAATCT 113488

161 uProProLysLysArgTyrTyrCysGluLeuCysLeuAspAspMetLeuT 178
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
113489 GCCGCCAAGAAGCGTTACTTTTTCGAACTGCAATTGGACAAGACGCTGT 113538

178 yrAlaArgThrThrSerLysProArgSerAlaSerGlyAspThrValphe 194
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
113539 ACGGCCGACTTCGGTGAAGCTGCAGACG.....GATCTGCTGTTT 113579

195 trpGlyGluHisPheGluPheAsnAsnLeuProAlaValArgAlaLeuAr 211
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
113580 TGGGGGAGCACATTCGATTTCCCGGACATACCCCGAGATTAATGTATCAC 113629

211 gLeuHisLeuTyrArgAspSerAspLysLysArgLysLysAspLysAlaG 228
::::: :||| :||| :||| :||| :||| :||| :||| :||| :|||
113630 TGTAAAGCTTTCCGTGAGGTGGACAAGAAGAAGACGGGACAAATACC 113679

228 lyTyrValGlyLeuValThrValProValAlaThrLeuAlaGlyArgHis 244
::::: :||| :||| :||| :||| :||| :||| :||| :||| :|||
113680 AATTTGGGATCCGTGAAGATACCCGTGCACGATGTCACTCCAGATTG 113729

```

245 PheThrGluGlnTrpTyrProValThrLeuProThrGlySerGlyGlySe 261  
|||||  
113730 CCCTGCGAACAATGGTATCCCAT...CTGAGCGACAAGCGGCGACAG 113776  
261 rGlyGlyMetGlySerGlyGlyGlySerGlySerGlySerGlyGly 278  
| ||| |||||  
113777 TCTGGCAGGACATCGGC.....GGCGGCGCAGTGGGT 113811  
278 lYsGlyLysGlyGlyCysProAlaValArgLeuLysAlaArgTyrGln 294  
: ||| ||| |||||  
113812 CCAAGGACAAGAGCAATTGCCACCGTGAGGATCAAGTGTCTTCCAG 113861  
295 ThrMetSerIleLeuProMetGluLeuTyrLysGluPheAlaGluTyrVa 311  
: : |||||  
113862 AGCACCGACATCTGCCCATCAATGTGTACGGCAACTTTTTCACGTACCT 113911  
311 lThrAsnHisTyrArgMetLeuCysAlaValLeuGluProAlaLeuAsnV 328  
: : : |||||  
113912 CAAGGAGAACATATAAGCGCGTGTGGAGACCCCTGGAGCGGTGATCGGAG 113961  
328 allYsGlyLysGluGluValAlaSerAlaLeuValHisIleLeuGlnSer 344  
|||||  
113962 TCAAGGCCAAGAGGACATTGGACAGGCACCTGGTGTCTGTATGCACGCA 114011  
345 ThrGlyLysAlaLysAspPheLeuSerAspMetAlaMetSerGluValAs 361  
||| ||| |||||  
114012 CAGGGATTGGCGGCGCCTTCTCTCACCGATGTGGTGGCCCTCGATCTGCT 114061  
361 pArgPheMetGluArgGluHisIlePheArgGluAsnThrLeuAlaT 378  
||| : : : ||| |||||  
114062 GCGAGTTGGCGAT..CAGAGGCTTACGTTTCAGGGGCAACTCCTTGGCCA 114108  
378 hrLysAlaIleGluGluTyrMetArgLeuIleGlyGlnLysTyrLeuLys 394  
|||||  
114109 CCAAGAGCATGGAGGCATTCTCTAAGCTGACGGCGGCAACAGTATCTGCAG 114158  
395 AspAlaIleGlyGluPheIleArgAlaLeuTyrGluSerGluGluAsnCy 411  
||| : : : ||| : : |||||  
114159 GACACACTATCCGCACCAATAAACGAGCTAATTTCAGTTCGGAGAGGGACTG 114208  
411 sGluValAspProIleLys.....CysThrAlaSerSerLeuAlaG 425  
|||||  
114209 CGAGGTGGATCCCAACCAAGACGAGCGGTTCTGTGCGCGGGTTCGCTGCAGC 114258  
425 luHisGlnAlaAsnLeuArgMetCysCysGluLeuAlaLeuCysLysVal 441  
: : ||||| : : : ||| : : :  
114259 GACAGCAGCGCGCTTGCCTGGCGCGGTCCGAGGGGCGGTGGCAGTGCATC 114308  
442 ValAsnSerHisCysValPheProArgGluLeuLysGluValPheAlase 458  
: : ||||| ||||| : : |||||  
114309 TTCGAATCGCACAAAGCATTTTCCCGCCCGATTACGAATAATGCTTTCGCAC 114358  
458 rTrpArgLeuArgCysAlaGluArgGlyArgGluAspIleAlaAspArgL 475  
: : ||| ||| : : |||||  
114359 GTTCCGGAGCGCTTGCAGCAGCTGGGCGGTTCAGGATATGGCCGACAACC 114408  
475 euIleSerAlaSerLeuPheLeuArgPheLeuCysProAlaIleMetSer 491  
|||||  
114409 TGATCTCGCGCAGCATTTCTCTGCGCTTCTGTGCGCGCATCTCTGTGCG 114458  
492 ProSerLeuPheGlyLeuMetGlnGluTyrProAspGluGlnThrSerAr 508  
|||||  
114459 CCGTCGCTCTCAATATCACCAGCAACTGCGCGTCCGCGACGGCCACCCG 114508  
508 qThrLeuThrIleAlaLysValIleGlnAsnLeuAlaAsnPheSerL 525  
| : : |||||  
114509 CAATCTCACACTGGTGGCCCAAGACCTGCAAAACATTGGCAATTTTCACCC 114558  
525 yspheThrSerLysGluAsppheLeuGlyPheMetAsnGluPheLeuGlu 541  
: ||| : : |||||  
114559 GCTTCCAGGGCAAGAGAACTTATGGAGTTTCTCAACGATTTCTCCGAG 114608  
542 LeuGluTrpGlySerMetGlnGlnPheLeuTyrGluIleSerAsnLeu.. 557

```

114609 CAGGAGCGCGCTCGCATGCAACAGTTTCTGGAGATTATATCCACGGGCC 114658
558 .....AspThrLeuThrAsnSerSerSerPheGluGlyT 569
114659 GGAGCACCCAGCTCCAGACTCGATCCTCGAT.....TGGSCCGGGT 114699
569 yrIleAspLeuGlyArgGluLeuSerThrLeuHisAlaLeuLeuTrpGlu 585
114700 ACATCGACAGGGCAACAGTTGTCCATACTACACAGTTTGTCTCAGCGAA 114749
586 ValLeuProGlnLeuSerLysGluAlaLeuLeuLysLeuGlyProLeuPr 602
114750 AGCCTGGCCAAGCTGCCGGAGGCCAGGCAGCAGAGCTGGATCCGTTGCA 114799
602 oArgLeuLeuSerAspIleSer.....ThrA 611
114800 GCATATTCTCGATGAATCAGCCGAGCCAGCAAGAGCATGGCATGGGCACAG 114849
611 laLeu..... 612
114850 CACTGCCGGGTGGATATTTCGCCGCCACCTCGTCTACGCACCTCGATAGCC 114899
613 .....ArgAsnPro..... 615
114900 AGCGAGAATCAGGAGAATCGCAATCCGGGATCCTCGGGCTCGGCACGCTGG 114949
616 ....AsnIleGlnArgGlnProSerArgGlnSerGluArgAlaArgSerG 631
114950 CTCCAACCTCGGAGCAGTTACTGCCACAACAAGCCAGTTGGCCCGAGCCGC 114999
631 InProMetValLeuArgGlyPro...SerAlaGlu.....MetGln 643
115000 AGCATGCGATTGTTAGTAACCATTTATCTCGGAGCGCGGCATCATCGCA 115049
644 Gly.TyrMetMetArgAspLeuAsnSerSerIleAspLeuGlnSerPheM 660
115050 GGAGTACTACGCCGAATCTCTGGAGAAGATATCTTTAGATACAAATGA 115099
660 etAlaArgGlyLeuAsnSerSerMetAspMetAlaArgLeuProSerPro 676
115100 TCCCACGGTTAATG..... 115113
677 ThrLysGluLysProProProProProGlyGlyGlyLysAspLeuPh 693
115113 ..... 115113
693 eTyrValSerArgProProLeuAlaArgSerSerProAlaTyrCysThrs 710
115114 .....GCTTACTGCAGCA 115126
710 erSerSerAspIleThrGluProGluGlnLysMetLeuSerValAsnLys 726
115127 GCAGCAGCAGCAGCAGCAGC.....AGCAGCAACAAC 115158
727 SerValSerMetLeuAspLeuGlnGlyAspGlyProGlyGlyArgLeuAs 743
115159 AGCAACAGCAACATC.....AGCAGCTGCAA 115184
743 nSerSerSerValSerAsn.....LeuAlaAlaValGly..AspLeuLe 757
115185 CAGCATGGCCATCAGCAACAGCCGACCCAGCAGCATCCACTCCAGATGCT 115234
757 uHisSerSerGlnAlaSerLeuThrAlaAlaLeuGlyLeuArgProAlaP 774
115235 CTCCAATTCAACAACCTCCATT...GCCGGCAACCAATATATGATTCGC 115281
774 roAlaGlyArgLeuSerGlnGlySerGlySerIleThrAlaAlaGly 790
115282 CAGGAGGCGCTGCAGCATGCCCAATCGCAGACCTCGATGGCGTCTCATCG 115331
791 MetArgLeuSerGlnMetGlyValThrThrAspGlyValProAlaGlnGI 807
:::

```

```

115332 CTT..... 115334
807 nLeuArgIleProLeuSerPheGlnAsnProLeuPheHisMetAlaAla 824
115334 ..... 115334
824 spGlyProGlyProAlaGlyHisGlyGlySerSerGly.....His 838
115335 .....AATGGGAGCAGCAGCAATTTGCTGCAC 115361
839 GlyProProSerSerHisHisHisHisHis.....Hi 851
115362 GGC.....CACCAGCAGCATGCCCATCACCCGAGCAACTGCA 115399
851 sHisHisArgGlyGlyGluProProGlyAspThrPheAlaPropPheHisG 868
115400 TCCACATCACTGC.....CGCCCG..... 115418
868 lyTyrSerLysSerGluAspLeuSerThrGlyValProLysProProAla 884
115419 .....GCGCCACAGACCAAGTGCC 115436
885 AlaSerIleLeuHis.....SerHisSerTyrSerAspGI 896
115437 TCCAGCACTATGGAGCGCATGGATCGCATGAACATATCCGTATATGTCGA 115486
896 uPheGlyProSerGlyThrAspPheThrArgArgGlnLeuSerLeuGlnA 913
115487 C.....AATGGCAATGACTACGAG..... 115505
913 spAsnLeuGlnHisMetLeuSerProGlnIleThrIleGlyProGln 929
115506 .....ACCAGCAGCGCTTCGAGCAGCTCGCTCCAGGACA 115538
930 ArgProAlaProSerGlyPro.....GlyGlyGlySerGlyGlyGlyse 944
115539 CTGCCACGGAATGGAAATCCCAATGCCAATGGCAACGTGGGAGCAGCAA 115588
944 rGlyGlyGlyGlyGlyGlnProProPro.LeuGlnArgGlyLysSer 960
115589 CAATAACCAAGCGCGCAGCTACGATGACATGCACGGGGAGTTCCAAATCC 115638
961 GlnGlnLeu.....ThrValSerAlaAlaGlnLysProAr 972
115639 AGATCTCTGGTTTCGATACGAGCAGTGTGTTTGTCTGCAAGTCGCCACA 115688
972 gPro.....Ser.SerGlyAsnLeuLeuGlnSerPro 982
115689 CCCATGATGAATCCAGTTTGGGACCAGCGGGGGCCGGACGAAGTCATCA 115738
983 GluProSerTyrGlyProAlaArgProArg..... 992
115739 CAAACTGAATTGGGAATACCCGATCACTCAGGTGGCTATGTGCGGGGTA 115788
993 .....GlnGlnSerLeuSerLysGluGlySerIleGlyGlySerGlyG 1007
115789 ATAATTGAATCCCAACTCGAATATGCCCAAGAACTTGGAGGATCTGGAC 115838
1007 lySer.....GlyGlyGlyGlyGlyGlyLeuLysPro 1018
115839 GATCTGTTCAAGTACCGCGAGGAGCATGACGTGGCGGAACCAAGCAACA 115888
1019 SerIleThrLysGlnHisSer...GlnThrProSerThr.LeuAsnPro. 1033
115889 TCACAATCACAAACCCAGGGTCAGCAGAGAACCAACAGGACATCTGAAGCCGG 115938
1034 ..ThrMetProAlaSerGluArgThrValAlaTrpValSerAsnMetPro 1049
115939 CCGCCGTTCGCGGAGGAGCAGCTGTCGGCGCAAAAGCAGT..... 115979
1050 HisLeuSerAlaAspIleGluSerAlaHisIleGluArgGluGluTyrLy 1066
115980 CACTGCAGTTCTGGCTACAGAGCATCTCCACAAATCCC..... 116018

```

1066 sLeuLysGluTyrSerLysSerMetAspGluSerArgLeuAspArgValL 1083  
116019 .....TCGCCCTCGCAGTCTCCAGTCCCGTGGAG..... 116048  
1083 ysGluTyrGluGluGluHisSerLeuLysGluArgLeuHisMetSer 1099  
116049 .....AGCCAGCTGAAGCGCGCGATGGCAGTCAC 116078  
11100 AsnArgLysLeuGluGluTyrGluArgArgLeuLeuSerGlnGluGlu 1116  
116079 AATGCGCGCGTG...GCCTTCAAGAAATCCCTCCTATCAGCTTCAGCCCCA 116125  
11116 nThrSerLysIleLeuMetGlnTyrGlnAlaArgLeuGluGlnSerGluL 1133  
116126 AACTGGCTCGTCCAGATCATCGGCACAGAGTAATACACACAGCAGCAGC 116175  
11133 ysArgLeuArgGlnGlnValGluLysAspSerGlnIleLysSerIle 1149  
116176 AGCAGCAGCAACAACAACAGTTT.....GGCAGCGCGCTTGAACCAATT 116219  
11150 IleGlyArgLeuMetLeuValGluGluGluLeuArgArgAspHisProAl 1166  
116220 GCGCGGCGACTGTFGCGCGCGAGGGCGGCTTCTCAACAGTGGCGGAGC 116269  
11166 aMetAlaGluProLeu.....ProGluProLys..... 1175  
116270 CTTGGAGGCGGCCACTTTGACGCCCGCCAGCTCCTCGGACGACAGCTCTCGG 116319  
11176 .....LysArgLeuLeuAspAlaGlnArgGlyse 1185  
116320 CGGATAATTACTTCAGTTATGCAGCGCGCTGCAGCTGTGGAGCGGTATT 116369  
11185 rPheProProTrrpValGlnGlnThrArgValArgTrrpProHisLeuGlyT 1202  
116370 GCCACCAAAATGSGAGGCTCAACGCTCG.....TCAGCGGCGGCAG 116410  
1202 hrAlaTrrpProGln..... 1207  
116411 TAGCTCTCCACCTCAGCATCTGCGTCCACCTCGAATCTTGGCAAGAGCG 116460  
1208 ...ProHisProProHisProGlyCysArgSerGlnArgThrAlaSerse 1223  
116461 GCGGCTCATCCGCTACGGCGGCTGAATGGCGCGCTGAAGCGCGGAGGAT 116510  
1223 rGlyThrProGlnThrThrSerProPro.....SerIleThrA 1236  
116511 GTCTACGGCAGTGCTACGGCGGCGCAGAGTGGAATGTGGCTATGGCTT 116560  
1236 spLeuLeuPro.Cys.....AlaProTyrProGlyProProSerValth 1250  
116561 GTCCACTTCCAGTGCCTGCGGCGGACACCATCAACATCCCCACAGCAGCAGC 116610  
1250 rAspLeuLeuProSerAlaProAspProGlyThrSerProThrThrArgT 1267  
116611 AGAATCCGATGCAGCAGCAGCAGCAAGAGGGAACGGGATCAGGAACAACAAG 116660  
1267 hr.....GlyArgHisGlnGlyThrAlaGlyLeuSerPro 1278  
116661 CAGTATGCCGCGCAGTGTGGCGGCGGCGGCGGATCGGCACATCAGCGGC 116710  
1279 LeuThrProPro.....TrpGlyThrHis 1286  
116711 TCAGAGCGCGCTGAGCTTGAGCTCGGCGCGCAC 116743  
seq\_name: gb\_htg4:AC012161  
seq\_documentation\_block:  
LOCUS AC012161 202741 bp DNA HTG 16-NOV-1999  
DEFINITION Drosophila melanogaster chromosome X clone BACR01122 (D1114)  
PCI-98 01.1.22 map 16B-16D strain y; cn bw sp, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 142 unordered pieces.  
ACCESSION AC012161

AC012161.7 GI:6435885  
HTG; HTGS\_PHASE1.  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 202741)  
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,  
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,  
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,  
Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,  
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.  
Sequencing of Drosophila melanogaster  
Unpublished  
2 (bases 1 to 202741)  
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,  
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,  
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.  
Direct Submission  
Submitted (21-OCT-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Nov 16, 1999 this sequence version replaced gi:6425635.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdgpefruitfly@berkeley.edu](mailto:bdgpefruitfly@berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 142 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 509: contig of 509 bp in length  
\* 510 589: gap of unknown length  
\* 590 1201: contig of 612 bp in length  
\* 1202 1281: gap of unknown length  
\* 1282 2394: contig of 1113 bp in length  
\* 2395 2474: gap of unknown length  
\* 2475 3512: contig of 1038 bp in length  
\* 3513 3592: gap of unknown length  
\* 3593 4500: contig of 908 bp in length  
\* 4501 4580: gap of unknown length  
\* 4581 5465: contig of 885 bp in length  
\* 5466 5545: gap of unknown length  
\* 5546 6253: contig of 708 bp in length  
\* 6254 6334: gap of unknown length  
\* 6334 7105: contig of 772 bp in length  
\* 7106 7185: gap of unknown length  
\* 7186 8077: contig of 892 bp in length  
\* 8078 8157: gap of unknown length  
\* 8158 9047: contig of 890 bp in length  
\* 9048 9127: gap of unknown length  
\* 9128 9530: contig of 403 bp in length  
\* 9531 9610: gap of unknown length  
\* 9611 10173: contig of 563 bp in length  
\* 10174 10253: gap of unknown length  
\* 10254 10841: contig of 588 bp in length  
\* 10842 10921: gap of unknown length  
\* 10922 11807: contig of 886 bp in length  
\* 11808 11887: gap of unknown length





126689 TCGCACGAGAGTCTGCTGTCCAGTCCAGCGCTCATGTCCACCATAGGTGA 126640  
82 ..... 82  
126639 GTTCTGGCTCAATAATGGAGTTTGTGGCAAAACCCCTCAAACCATTTCTTTT 126590  
83 .....LeuGluLeuAsnLeuAspGluAspSerIleIleLysPro 95  
126589 TCTTTCTTCTAGATCTCTCGTGCACCTGGGCGGTGGCGGTGGCGCC 126540  
96 ValHisSerSerIleLeuGlyGlnGluPheCysPheGluValThrThrSe 112  
126539 GTGCATCAGTCGGTCTGGGACGTCGTCACCTGTTTCCAGGTACGGGCGG 126490  
112 r...SerGlyThrLysCysPheAlaCysArgSerAlaAlaGluArgAspL 128  
126489 GCCTCGTGGCGAGCGGTACTACTATCGCGGATCGCGCCAGGAGCGCGACC 126440  
128 ysTrpIleGluAsnLeuGlnArgAlaValLysProAsnLysAspAsnSer 144  
126439 TTTGGATCTACTCGCTGCGCAAGTCGATCGCTCCGAATCGAGAGCACACG 126390  
145 ArgArgValAspAsnValLeuLysLeuTrpIleIleGluAlaArgGluLe 161  
126389 CGTCGCACGGACAACCTCGCTGAAGATGTGGTGTACGAGCGGAAATCT 126340  
161 uProProLysLysArgTyrTyrCysGluLeuCysLeuAspMetLeu 178  
126339 GCCGCCCAAGAGCGTTACTTTTGGCAACTGCAATTTGGACAAGACGCTGT 126290  
178 yrAlaArgThrThrSerLysProArgSerAlaSerGlyAspThrValphe 194  
126289 ACGCGCGGACTTCGGTGAAGCTGCAGACG.....GATCTGCTGTTT 126249  
195 TrpGlyGluHisPheGluPheAsnAsnLeuProAlaValArgAlaLeuAr 211  
126248 TGGGGGAGCACTTCGATTTCCCGACATACCCGAGATTAATGTGATCAC 126199  
211 gLeuHisLeuThrArgAspSerAspLysLysArgLysLysAspLysAlag 228  
126198 TGTTAACGTTTTCGCTGAGGTGGACAAGAAGAAGAGCGGACAAATACC 126149  
228 lyTyrValGlyLeuValThrValProValAlaThrLeuAlaGlyArgHis 244  
126148 AATTTGTGGATCGGTGAAGATACCCGTCACGATGTCACTCCAGATTG 126099  
245 PheThrGluGlnTrpTyrProValThrLeuProThrGlySerGlyGlyse 261  
126098 CCCTGCGAACAATGGTATCCCAT...CTGAGCGACAAGCGGGCGACAG 126052  
261 rGlyGlyMetGlySerGlyGlyGlyGlySerGlyGlySerGlyGly 278  
126051 TCTGGGCAGGACATCGGGC.....GGCGGGCGCAGTGGGT 126017  
278 lyLysGlyLysGlyGlyCysProAlaValArgLeuLysAlaArgTyrGln 294  
126016 CCAAGGACAAGGAGCAATTTGCCACGCTGAGGATCAAGTGTCTTCCAG 125967  
295 ThrMetSerIleLeuProMetGluLeuTyrLysGluPheAlaGluTyrVa 311  
125966 AGCACCGACATCTCGTCCCATCAATGTGTACGGCAACTTTTGTACGTACCT 125917  
311 lThrAsnHistyArgMetLeuCysAlaValLeuGluProAlaLeuAsnV 328  
125916 CAAGGAGAACTATAAGCGCGTGTGCGAGACCCCTGGAGCGCGGTGATCGGAG 125867  
328 alLysGlyLysGluGluValAlaSerAlaLeuValHisIleLeuGlnSer 344  
125866 TCAAGGCCAAGGAGGACATTGGACAGCACTGGTGTGCTGTGATGACGCA 125817  
345 ThrGlyLysAlaLysAspPheLeuSerAspMetAlaMetSerGluValAs 361  
125817

125816 CAGGGATTGGCGGGCGCCTTCTCACCAGATGTGGTGGCCCTCGATCTGCT 125767  
361 pArgPheMetGluArgGluHisLeuIlePheArgGluAsnThrLeuAlaT 378  
125766 GCGAGTTGGCGAT...CAGAGGCTTACGTTACAGGGGCAACTCCTTGGCCA 125720  
378 hrLysAlaIleGluGluTyrMetArgLeuIleGlyGlnLysTyrLeuLys 394  
125719 CCAAGAGCATGGAGGCATTCCTCAAGCTGACGGGCGCAACAGTATCTGCAG 125670  
395 AspAlaIleGlyGluPheIleArgAlaLeuTyrGluSerGluGluAsnCy 411  
125669 GACACACTATCCGCACCAATAAACGAGCTAATTCAGTCGGAGAGGACTG 125620  
411 sGluValAspProIleLys.....CysThrAlaSerSerLeuAlaG 425  
125619 CGAGGTGGATCCCAACAGACGAGCGGTTCGTCGGCGGTTTCGCTGCAGC 125570  
425 luHisGlnAlaAsnLeuArgMetCysCysGluLeuAlaLeuCysLysVal 441  
125569 GACAGCAGCGCGCTTCGTCGGCGGCTCCGAGGGCGTGGCAGTGCATC 125520  
442 ValAsnSerHisCysValPheProArgGluLeuLysGluValPheAlase 458  
125519 TTCGAAATCGACAAAGCATTTTCCCGCCCGAGTTACGAAATTCGCTTTCGAC 125470  
458 rTrpArgLeuArgCysAlaGluArgGlyArgGluAspIleAlaAspArgL 475  
125469 GTTCGGGAGCGCTTCAGCAGCAGCTGGGCGCTCAGGATATGGCCGACACC 125420  
475 euIleSerAlaSerLeuPheLeuArgPheLeuCysProAlaIleMetSer 491  
125419 TGATCTCGCGAGCATTTTCTCGCTTCCTGTGCCCGGCCATCTCTGTCG 125370  
492 ProSerLeuPheGlyLeuMetGlnGluTyrProAspGluGlnThrSerAr 508  
125369 CCGTCGCTCTTCAATATCACCAGCACTCCCGTCCGCGCCGACCCG 125320  
508 gThrLeuThrLeuIleAlaLysValIleGlnAsnLeuAlaAsnPheSerL 525  
125319 CAATCTCACACTGGTGGCCCAAGACCCCTGCAAAACATTGGCCCAATTCACCC 125270  
525 yspheThrSerLysGluAspPheLeuGlyPheMetAsnGluPheLeuGlu 541  
125269 GCTTCCAGGGCAAGAGAACTTTATGGAGTTTCTCAACGATTCTCTCGAG 125220  
542 LeuGluTrpGlySerMetGlnGlnPheLeuTyrGluIleSerAsnLeu... 557  
125219 CAGGAGCGCGCTCGCATGCAACAGTTTCTGGAGATTATATCCAGCGGGCC 125170  
558 .....AspThrLeuThrAsnSerSerSerPheGluGlyT 569  
125169 GGAGCACCCAGCTCCAGACTCGATCTCGAT.....TGGGCGGGGT 125129  
569 yrIleAspLeuGlyArgGluLeuSerThrLeuHisAlaLeuLeuTrpGlu 585  
125128 ACATCGACCCAGGGCAACAGTTGTCCATACTACACAGTTTGTCTCAGCGAA 125079  
586 ValLeuProGlnLeuSerLysGluAlaLeuLysLeuGlyProLeuPr 602  
125078 AGCCTGGCCAAGCTGCCGGAGGGCCAGGACGACGAGCTGGATCCGTTGCA 125029  
602 oArgLeuLeuSerAspIleSer.....ThrA 611  
125028 GCATATTCTCGATGAATCAGCCGAGGCCAAAGAGCATGGCATGGGCACAG 124979  
611 laLeu..... 612  
124978 CACTGCCGGGTGGATATTGGCCGGCCACCTCGTCTACGCACCTCGATAGCC 124929  
613 .....ArgAsnPro..... 615  
124928 AGCGAGAATCAGGAGAATCGCAATCCGGGATCCTCGGGCTCGCACGCTGG 124879

616 ....AsnIleGlnArgGlnProSerArgGlnSerGluArgAlaArgSerG 631  
124878 CTCCAACCTGGAGCAGTACTGCCACAAAGCCAGTGGCCCGAGCCGC 124829  
631 lnProMetValLeuArgGlyPro...SerAlaGlu.....MetGln 643  
124828 AGCATGCGATTGTTAGTAAACCATTTATCTGGGAGCGCGGCATCATCGGA 124779  
644 Gly.TyrMetMetArgAspLeuAsnSerSerIleAspLeuGlnSerPheM 660  
124778 GGAGTACTTACCGCGAATTCTCTGGAGAGAAATATCTTTAGATACAATGA 124729  
660 etAlaArgGlyLeuAsnSerSerMetAspMetAlaArgLeuProSerPro 676  
124728 TCCACCGTTAATG..... 124715  
677 ThrLysGluLysProProProProProGlyGlyGlyLysAspLeuPh 693  
124715 ..... 124715  
693 eTyrValSerArgProProLeuAlaArgSerSerProAlaTyrCysThrs 710  
124714 .....GCTTACTGCAGCA 124702  
710 erSerSerAspIleThrGluProGluGlnLysMetLeuSerValAsnLys 726  
124701 GCAGCAGCAGCAGCAGCAGC.....AGCAGCAACAAC 124670  
727 SerValSerMetLeuAspLeuGlnGlyAspGlyProGlyGlyArgLeuAs 743  
124669 AGCAACAGCAACATC.....AGCAGCTGCAA 124644  
743 nSerSerSerValSerAsn.....LeuAlaAlaValGly..AspLeuLe 757  
124643 CAGCATGGCCATCAGCAACAGCCGCACACCAGCATCCACTCCAGATGCT 124594  
757 uHisSerSerGlnAlaSerLeuThrAlaAlaLeuGlyLeuArgProAlaP 774  
124593 CTCCAATTCAACAACCTCCATT...GCCGGACCAATATATGAGTTCGC 124547  
774 roAlaGlyArgLeuSerGlnGlySerGlySerIleThrAlaAlaGly 790  
124546 CAGGAGGCTGCAGCATGCCCAATCGCAGACCTCGATGGCTCCTCATCG 124497  
791 MetArgLeuSerGlnMetGlyValThrThrAspGlyValProAlaGlnG 807  
124496 CTT..... 124494  
807 nLeuArgIleProLeuSerPheGlnAsnProLeuPheHisMetAlaAla 824  
124494 ..... 124494  
824 spGlyProGlyProProAlaGlyHisGlyGlySerSerGly.....His 838  
124493 .....AATGGGAGCAGCAGCAATTGCTGCAC 124467  
839 GlyProProSerSerHisHisHisHisHisHis.....H1 851  
124466 GGC.....CACCAGCAGCATGCCCATCACCCCGCAGCAACTGCA 124429  
851 sHisHisArgGlyGlyGluProProGlyAspThrPheAlaPropHeHisG 868  
124428 TCCACATCACTGC.....CGCGCG..... 124410  
868 lyTyrSerLysSerGluAspLeuSerThrGlyValProLysProProAla 884  
124409 .....GCGCCACAGACCAGTGCC 124392  
885 AlaSerIleLeuHis.....SerHisSerTyrSerAspG 896  
124391 TCCAGCACTATGGAGCGCATGGATCGCATGAACATATCCGTATATGTCGCA 124342

896 uPheGlyProSerGlyThrAspPheThrArgArgGlnLeuSerLeuGlnA 913  
124341 C.....AATGGCAATGACTACGAG..... 124323  
913 spAsnLeuGlnHisMetLeuSerProProGlnIleThrIleGlyProGln 929  
124322 .....ACCAGCAGCCTTCGAGCAGCTCGCTCCAGGACA 124290  
930 ArgProAlaProSerGlyPro.....GlyGlyGlySerGlyGlySe 944  
124289 CTGCCACGGAATGGAATCCCAATGCCAATGGCAACGTGGGCGAGCA 124240  
944 rGlyGlyGlyGlyGlyGlnProProProLeuGlnArgGlyLysSer 960  
124239 CAATAACCAGAGCGGCGAGCTACGATGACATGCACGGGGAGTTCCAAATCC 124190  
961 GlnGlnLeu.....ThrValSerAlaAlaGlnLysProAr 972  
124189 AGATCTCTGGGTTCGATACGAGCAGTGTCTTTGTCTGCAAGTCGCCACA 124140  
972 qPro.....Ser.SerGlyAsnLeuLeuGlnSerPro 982  
124139 CCCATGATGAATCCAGTTTGGGACACGCGGGCGCGGACCAAGTCATCA 124090  
983 GluProSerTyrGlyProAlaArgProArg..... 992  
124089 CAAACTGAATTTGGGAATACCCGATCACTCAGGTGGCTATGTGCGGGGTA 124040  
993 .....GlnGlnSerLeuSerLysGluGlySerIleGlyGlySerGlyG 1007  
124039 ATAATTTGAATCCCAACTCGAATATGCCCAAGAACTTGGAGGATCTGGAC 123990  
1007 lySer.....GlyGlyGlyGlyGlyGlyLeuLysPro 1018  
123989 GATCTGTTCAAGTACGCCGAGGAGCATACGTGGCGGAACCAAGCAACCA 123940  
1019 SerIleThrLysGlnHisSer...GlnThrProSerThr.LeuAsnPro. 1033  
123939 TCACAATCACAAACCAGGGTCAGCAGAACCAACAGGACATCTGAAGCCGG 123890  
1034 ..ThrMetProAlaSerGluArgThrValAlaTrpValSerAsnMetPro 1049  
123889 CCGCCGTTCCCGGCAAGGAGCAGCTGTCGGCGAAAGCAGT..... 123849  
1050 HisLeuSerAlaAspIleGluSerAlaHisIleGluArgGluGluTyrLy 1066  
123848 CACTGCAGTTCTGGCTACCAGAGCATCTCCACAATCCC..... 123810  
1066 sleuLysGluTyrSerLysSerMetAspGluSerArgLeuAspArgValL 1083  
123809 .....TCGCCCTCGCAGTCTCTCAGTCCCGTGGAG..... 123780  
1083 ysGluTyrGluGluGluIleHisSerLeuLysGluArgLeuHisMetSer 1099  
123779 .....AGCCAGCTGAAGCCCGCGATGGCAGTCAC 123750  
1100 AsnArgLysLeuGluGluTyrGluArgArgLeuLeuSerGlnGluGlu 1116  
123749 AATGCGCGCTG...GCCTTCAAGAAATCCCTCTATCAGCTTCAGCCCCA 123703  
1116 nThrSerLysIleLeuMetGlnTyrGlnAlaArgLeuGluGlnSerGluL 1133  
123702 AACTGGCTCGTCCAGATCATCGGCACAGTAATAACACACAGCAGC 123653  
1133 ysArgLeuArgGlnGlnValGluLysAspSerGlnIleLysSerIle 1149  
123652 AGCAGCAGCAACAACAACAGTTT.....GGCAGCGCTTGAACCAAT 123609  
1150 IleGlyArgLeuMetLeuValGluGluArgArgAspHisProAl 1166  
123608 GCGCGGAGTGGTGGCGCGGAGGCGGCTTCTCAACAGTGGCGGAGC 123559  
1166 aMetAlaGluProLeu.....ProGluProLys..... 1175



```
123558 CTGGAGCGGCCACCTTTGACGCCCTCCTCGGACGACAGCTCTCGG 123509
1176 .....LysArgLeuLeuAspAlaGlnArgGlySe 1185
123508 CGGATAATTACTTCAGTTATGACGGCTGCGAGCTGCTGGAGCGGTATT 123459
1185 rPheProProTrpValGlnGlnThrArgValArgTrpProHisLeuGlyT 1202
123458 GCCACCAATTTGGAGCTCAACGCTCGC.....TCACGGCGGCAG 123418
1202 hrAlaTrpProProGln..... 1207
123417 TAGCTCCTCCACCTCAGCATCTGCGTCCACCTCGAATCTTGGCAAGAGCG 123368
1208 ....ProHisProProHisProGlyCysArgSerGlnArgThrAlaSerSe 1223
123367 GCGGCTCATCCGCTACGGCGGCTGATGGCCGCTGAAGCGCGAGGAT 123318
1223 rGlyThrProGlnThrThrSerProPro.....SerIleThrA 1236
123317 GTCTACGGCAGTGGCTACGGCGGCGAGCAGTGGAAATGTGGGCTATGGCTT 123268
1236 sPLeuLeuPro.Cys.....AlaProTrpProGlyProProSerValTh 1250
123267 GTCCACTTCCAGTGGCGGGGACACCATCAACATCCCCACGACGAGCAGC 123218
1250 rAspLeuLeuProSerAlaProAspProGlyThrSerProThrThrArgT 1267
123217 AGAATCCGATGCAGCAGCAGCAGCAAGGGAACGGGATCAGGAACACAAG 123168
1267 hr.....GlyArgHisGlnGlyThrAlaGlyLeuSerPro 1278
123167 CAGTATGCCGCGAGTGGCGGGCAGCGTGGGATCGGCCATCAGCGGC 123118
1279 LeuThrProPro.....TrpGlyThrHis 1286
123117 TCAGAGCGCCTGAGCTTGAGTGGAGTGGCGGCGCAC 123085
```

seq\_name: gb\_inl:AE003506

```
seq_documentation_block:
LOCUS AE003506 300994 bp DNA INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386053 section 23
of 30, complete sequence.
ACCESSION AE003506 AE002593
VERSION AE003506.1 GI:7293355
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 300994)
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
```

```
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Shue,B.C., Siden-Kiamos,I.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Spradling,A.C.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Tector,C., Turner,R.,
Stapleton,M., Strong,R., Sun,E., Svirska,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
2 (bases 1 to 300994)
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
FEATURES
Location/Qualifiers
1..300994
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="X"
complement(join(<16198..16317,16386..16574,16635..16845,
16915..17053,17129..17411,17526..17607,17676..17975,
18107..>18483))
/gene="CG8527"
/product="CT24901"
/db_xref="FLYBASE:FBan0008527"
/db_xref="FLYBASE:FBgn0030844"
complement(join(16198..16317,16386..16574,16635..16845,
16915..17053,17129..17411,17526..17607,17676..17975,
18107..18483))
/gene="CG8527"
/note="CG8527 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0008527"
/db_xref="FLYBASE:FBgn0030844"
/protein_id="AAF48734.1"
/db_xref="GI:7293356"
/translation="MPQEKRRHPPESTRQRELETLVIFRRSLIYQTKFQNSTLHGVR
YIAESGRPIGEKFMWFCFTSIGAVTALVIMSLWEKFQTNPTIGLDTDFHNQNVFPP
TTVVCPEAAFDHDKTYEKVYNTLANVDEAQAQMYTPFLRILTSLNFNENVRDAKVLQS
IPQNLDAHTIREWAFEGHIDCKNVFVSCYRDEDIPCCDHEPIYTEHGFCYAFNSR
FKSTPTEDVKTGAPHDLYETDKKWFALFPNSTRIFISNEYFGSDFNQIDWSEP
QLVEVRISKNTYTTDDARQLSIGQRKCIFSDVKLVNFPDAYTFSSCMKQCRMNAI
KLCKCNPPFYKPIPNVPMCSIKDFCDLDEFKSNITNICKLOCELSCKTTFNIDKLI
KNSDRPESGLVLELTWPIIRYKREVLFGWVDLLVSFGGIALSLFLGSLLSGVEIY
YFTLRACCMVMYKRNQLEYIEEKIRQEPKIDKLKLSKNPRIGDTPTSAVLKVP
AEESAGPNPLGGYQRKHEKDYSLNTKNYKTPKVPVDYGYTSKTKNSWTYDHSQYMP
"
complement(join(21273..21792,21868..21999,22058..22191,
22257..22823,22889..23887,24168..24320,24398..25116,
25191..25449))
/gene="CG8465"
/product="CT24743"
/db_xref="FLYBASE:FBan0008465"
/db_xref="FLYBASE:FBgn0030845"
complement(join(21273..21792,21868..21999,22058..22191,
```



```
161 uProProLysLysArgTyrTyrCysGluLeuLeuCysLeuAspMetLeu 178
|||||
196267 GCCGCCAAGAAGCGTTACTTTGCGAAGTGAATGGACAAGACGCTGT 196316
178 YrAlaArgThrThrSerLysProArgSerAlaSerGlyAspThrValPhe 194
|||||
196317 ACGCCGGACTTCGGTGAAGCTGCAGAGC.....GATCGCTGTTT 196357
195 TrpGlyGluHisPheGluPheAsnAsnLeuProAlaValArgAlaLeuAr 211
|||||
196358 TGGGGGGAGCATTTCGATTTCGCCGACATACCCGAGATTAAATGTGATCAC 196407
211 gLeuHisLeuTyrArgAspSerAspLysLysArgLysLysAspLysAlaG 228
|||||
196408 TGTAAACGTTTTCGGTGAGGTGGACAAAGAAGAACGGGACAAATACC 196457
228 lyTyrValGlyLeuValThrValProValAlaThrLeuAlaGlyArgHis 244
|||||
196458 AATTGTGGGATCGGTGAAGATACCCGTCACGATGTCACTCCAGATTG 196507
245 PheThrGluGlnTrpTyrProValThrLeuProThrGlySerGlyGlyse 261
|||||
196508 CCCTGCGAACAATGTTATCCCAT...CTGAGCGACAAGCGGGCGACAG 196554
261 rGlyGlyMetGlySerGlyGlyGlyGlyGlySerGlyGlyGlySerGlyG 278
|||||
196555 TCTGGGCAGGACATCGGGC.....GGCGGCGCAGTGGGT 196589
278 lyLysGlyLysGlyGlyCysProAlaValArgLeuLysAlaArgTyrGln 294
|||||
196590 CCAAGGACAAGGACCAATTTGCCACCGTGAGGATCAAGTGTCTGTTCCAG 196639
295 ThrMetSerIleLeuProMetGluLeuTyrLysGluPheAlaGluTyrVa 311
|||||
196640 AGCACCGACATCTGCCCATCAATGTACGGCAACTTTTGTGACGTACCT 196689
311 lThrAsnHisTyrArgMetLeuCysAlaValLeuGluProAlaLeuAsnV 328
|||||
196690 CAAGGAGAACTATAAGCGCGTGTGCGAGACCCCTGGAGCGCGTGTGCGAG 196739
328 alLysGlyLysGluGluValAlaSerAlaLeuValHisIleLeuGlnSer 344
|||||
196740 TCAAGGCCAAGGAGGACATTTGGACAGCAGTGTGCTGTGATGCAGCA 196789
345 ThrGlyLysAlaLysAspPheLeuSerAspMetAlaMetSerGluValAs 361
|||||
196790 CAGGGATTGGCGGCGCTTCTCACCAGATGGTGGCCCTCGATCTGCT 196839
361 pArgPheMetGluArgGluHisLeuPheArgGluAsnThrLeuAlat 378
|||||
196840 GCGAGTTGGCGAT...CAGAGGCTTACGTTACGGGGCAACTCTTGCCCA 196886
378 hrLysAlaIleGluGluTyrMetArgLeuIleGlyGlnLysTyrLeuLys 394
|||||
196887 CCAAGAGCATGGAGGCATTCTCAAGCTGACGGGGCAACAGTATCTGCAG 196936
395 AspAlaIleGlyGluPheIleArgAlaLeuTyrGluSerGluGluAsnCy 411
|||||
196937 GACACACTATCCGCACCAATAAACGAGCTTAATTCACTCGGAGAGGGACTG 196986
411 sGluValAspProIleLys.....CysThrAlaSerSerLeuAlaG 425
|||||
196987 CGAGGTGGATCCACCAAGACGAGCGGTTCGTCGGCGGTTTCGCTGCAGC 197036
425 luHisGlnAlaAsnLeuArgMetCysGluLeuAlaLeuCysLysVal 441
|||||
197037 GACAGCGCGCGCTTGGTGGCGGCTCCGAGGGCGGTGGCAGTGCATC 197086
442 ValAsnSerHisCysValPheProArgGluLeuLysGluValPheAlaSe 458
|||||
197087 TTCGAATCGCACAAAGCATTTTCCCGCCAGTTACGAAATGCTTTGCGAC 197136
458 rTrpArgLeuArgCysAlaGluArgGlyArgGluAspIleAlaAspArgL 475
|||||
197137 GTTCGGGAGCGCTTGCAGCAGCTGGCGCGTACGATATGGCCGACAACC 197186
475 euIleSerAlaSerLeuPheLeuArgPheLeuCysProAlaIleMetSer 491
|||||
197187 TGATCTCGGCGAGCATTTCTCGCTTCTCTGCGCTTCTCTGCGGCAATCTGTG 197236
492 ProSerLeuPheGlyLeuMetGlnGluTyrProAspGluGlnThrSerAr 508
|||||
197237 CCGTCGCTCTTCAATATCACAGCGAACTGCGGTCCGCACGGGGCACCCG 197286
508 gThrLeuThrLeuIleAlaLysValIleGlnAsnLeuAlaAsnPheSerL 525
|||||
197287 CAATCTCACACTGGTGGCCAAAGACCTTGCAAAACATTTGCCAATTTCACCC 197336
525 yspheThrSerLysGluAspPheLeuGlyPheMetAsnGluPheLeuGlu 541
|||||
197337 GCTTCCAGGGCAAGAGAACTTTATGGAGTTTCTCAACGATTTCCTCGAG 197386
542 LeuGluTrpGlySerMetGlnGlnPheLeuTyrGluIleSerAsnLeu... 557
|||||
197387 CAGGAGCGCTCGCATGCAACAGTTTCTGGAGATTATATCCACCGCGGCC 197436
558 .....AspThrLeuThrAsnSerSerPheGluGlyT 569
|||||
197437 GGAGCACCCAGCTCCAGACTCGATCCTCGAT.....TGGCGCGGT 197477
569 yrIleAspLeuGlyArgGluLeuSerThrLeuHisAlaLeuLeuTrpGlu 585
|||||
197478 ACATCGACCCAGGGCAACAGTTGTCCATCTACACAGTTTGTCTACCGAA 197527
586 ValLeuProGlnLeuSerLysGluAlaLeuLeuLysLeuGlyProLeuPr 602
|||||
197528 AGCCTGGCCAAGCTGCGGAGGCCAGGCGACGACGAGTGGATCCGTTGCA 197577
602 oArgLeuLeuSerAspIleSer.....ThrA 611
|||||
197578 GCATATCTCGATGAATCAGCCGAGGCCAAGAGCATGGCATGGGCACAG 197627
611 laLeu..... 612
|||||
197628 CACTGCCGGTGGATATTGCGGGCCACCTCGTCTACGCACCTCGATAGCC 197677
613 .....ArgAsnPro..... 615
|||||
197678 ACCGAGAATCAGGAGAAATCGCAATCCGGGATCTCGGGCTCGCACGCTGG 197727
616 ....AsnIleGlnArgGlnProSerArgGlnSerGluArgAlaArgSerG 631
|||||
197728 CTCCAACCTCGGAGCAGTTACTGCCACACAACAAAGCCAGTTGGCCCGCCG 197777
631 lnProMetValLeuArgGlyPro...SerAlaGlu.....MetGln 643
|||||
197778 ACCATGCGATTGTTAGTAAACCATTTATCTCGGAGCGCGGCATCATGCGA 197827
644 Gly.TyrMetMetArgAspLeuAsnSerSerIleAspLeuGlnSerPheM 660
|||||
197828 GGAGTACTTACGCCGGAATTCTCTGGAGAGAATAATCTTTAGATACAATGA 197877
660 etAlaArgGlyLeuAsnSerSerMetAspMetAlaArgLeuProSerPro 676
|||||
197878 TCCCACGGTTAATG..... 197891
677 ThrLysGluLysProProProProProGlyGlyLysAspLeuPh 693
|||||
197891 ..... 197891
693 eTyrValSerArgProProLeuAlaArgSerSerProAlaTyrCysThrS 710
|||||
197892 .....GCTTACTGCAGCA 197904
710 erSerSerAspIleThrGluProGluGlnLysMetLeuSerValAsnLys 726
```



[illegible]







```

:::
2847 CATCT..... 2851

803 ValProAlaGlnGlnLeuArgIleProLeuSerPheGlnAsnProLeuPh 819
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2852 ...CCAGTTGAACGAATG...GCCGCTCTATCAATGCTAACCCATGCTT 2895

819 eHisMetAlaAlaAspGlyProGlyProGlyProAlaGlyHisGly..... 833
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2896 T.....GGACAGGCCCATCATCTGGATATGCTATACCTG 2930

834 .....GlySerSerGlyHisGlyProProSer 842
::: |||:
2931 CAGAGCCAAAGGAATCGTATACCAAAGCGAGCAAGCCACCATAC 2980

843 SerHisHisHisHisHisHisHisHis 851
::: |||:||||:||||:
2981 GATCCCGATGTGCACAACTATCATTTAT 3007

seq_name: gb_pr4:AK022662

seq_documentation_block: 2385 bp mRNA PRI 29-SEP-2000
LOCUS AK022662 Homo sapiens cDNA FLJ12600 fis, clone NT2RM4001412, highly similar
DEFINITION to Homo sapiens ngap mRNA.
ACCESSION AK022662
VERSION AK022662.1 GI:10434186
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,
clone_lib:NT2RM4 clone:NT2RM4001412.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2385)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
1..2385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone="NT2RM4001412"
/clone_lib="NT2RM4"
/note="cloning vector: pME18SFL3-mRNA from uninduced NT2
neuronal precursor cells."
BASE COUNT 690 a 602 c 576 g 517 t
ORIGIN
alignment_scores: Quality: 1088.50 Length: 878

```

```

Ratio: 2.143 Gaps: 39
Percent Similarity: 57.859 Percent Identity: 36.788
alignment_block:
US-09-294-298-6 x AK022662 ..
Align seg 1/1 to: AK022662 from: 1 to: 2385
454 GluValPheAlaSerTrpArgLeuArgCysAlaGluArgGlyArgGluAs 470
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2 GAAAGTGTTCATCATCATGGAAGCAGCAGTGCCTGAACCGTGGCAAGCAAGA 51
470 PileAlaAspArgLeuIleSerAlaSerLeuPheLeuArgPheLeuCysP 487
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
52 CATCAGCGAGAGGCTCATCAGTGCCTCATATTTCTCCGTTTCTGTGTC 101
487 roAlaIleMetSerProSerLeuPheGlyLeuMetGlnGluTyrProAsp 503
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
102 CAGCCATTATGTCTCCAGTCTTTCAACCTTATGCAGGAGTATCTTGAT 151
504 GluGlnThrSerArgThrLeuThrLeuIleAlaLysValIleGlnAsnLe 520
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
152 GACCGCACATCTCGGACTCTAACTCTTATGCCAAGGTCAATCAGAACCT 201
520 uAlaAsnPheSerLysPheThrSerLysGluAspPheLeuGlyPheMetA 537
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
202 GGCCAACTTTGCCAAGTTGGTAACAAAGAGGAATACATGGCATTCTATGA 251
537 snGluPheLeuGluLeuGluTrpGlySerMetGlnGlnPheLeuTyrGlu 553
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
252 ATGATTTTGTAGAACATGAATGGGGTGGATGAAGCGCTTTCTTTGGAG 301
554 IleSerAsnLeuAspThrLeuThrAsnSerSerPheGluGlyTyrIle 570
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
302 ATCTCTAATCCAGACACCATCTCAAAACACCCAGGCTTTGATGGTTACAT 351
570 eAspLeuGlyArgGluLeuSerThrLeuHisAlaLeuLeuTrpGluValL 587
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
352 TGATCTGGCGGAGAGCTTTCAGTTTGGCATTCCTTACTGTGGGAAGTAG 401
587 euProGlnLeuSerLys.....GluAlaLeuLeu 596
::: |||:||||:||||:
402 TTTCCCAACTTGATAAGGGTGAAATTCCTCTACAGGCGCGCTGGCA 451
597 LysLeuGlyProLeuProArgLeuLeuSerAspIleSerThrAlaLeuAr 613
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
452 AAATGGGGCCTCTCCCTCGTGTCTTGTGTGATATTACCAAGTCATTGAC 501
613 gAsnProAsn...IleGlnArgGlnProSerArgGlnSerGluArgAlaA 629
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
502 TAATCTACGCCCAATACAAACAGCAACTGACAGCGCTTCACTGAACAT... 547
629 rgSerGlnProMetValLeuArgGlyProSerAlaGluMetGlnGlyTyr 645
547 ..... 547
646 MetMetArgAspLeuAsnSerSerIleAspLeuGlnSerPheMetAlaAr 662
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
548 .....AACTCCAGTCCAAATGTCAAGTGGAGCGCTCTCTCTC 582
662 gGlyLeuAsn.....SerSerMetAspMetAlaArgL 673
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
583 TGGGCTGCAGAAATAATTGAAGACCCCACTGACAGTATTGTCATAAAC 632
673 euProSerProThrLysGluLysProProProProProGlyGlyGly 689
|| | ||| |||:||||:||||:
633 TAAATCTCCAAGCCAGGACAAC.....ACAGACAGCTACTTCAGAGGG 676
690 LysAspLeuPheTyrValSerArgProProLeuAlaArgSerSerProAl 706
|| | ||| |||:||||:||||:
677 AAAACATTATTGCTGTTTCAGCAA.....GCCTCTCTCTCAGAG 714
706 atYrCysThrSerSerSerAspIleThrGluProGluGlnLysMetLeuS 723

```

```

1390 GAATGGGAGCGGTCCCGCAGCAGTCTCTTCTTCCAGAGAGAGCCCTG 1439
1002 leGlyGlySerGlyGlySerGlyGlyGlyGlyGlyGlyGlyGlyLeuLysPro 1018
:: :|||
1440 TT..... 1444
1019 SerIleThrLys...GlnHisSerGlnThrProSerThrLeuAsnPro... 1033
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1445 AAAGTTAGAGCAATCCAGAGACAACAGACACAGCAGCGTTCAGTCACTGT 1494
1034 .....ThrMetProAlaSerGluArgThrValAlaTrpValSerA 1047
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1495 GGACTCTGCCACAATGTCCCAGTAGAGAGGACAGCAGCCTGGTTCGA 1544
1047 snMetProHisLeuSerAlaAspIleGluSerAlaHisIleGluArgGlu 1063
|| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1545 ACAATGGCAGTATGAAGAGGATGTGAA..... 1573
1064 GluTyrLysLeuLysGluTyrSerLysSerMetAspGluSerArgLeuAs 1080
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1574 .....GAACCTGAGCAAAATCTAGATGAAGCCAAAG..... 1603
1080 pArgValLysGluTyrGluGluGluIleHisSerLeuLysGluArgLeuH 1097
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1604 .CATGCTGAGAAGTATGAACAAGAAATTAATAACTGAAGGAGCGCCTGA 1652
1097 isMetSerAsnArgLysLeuGluGluTyrGluArgArgLeuLeuSerGln 1113
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1653 GAGTTCCAGCGCGGACTGGAGGAATATGAACGCCGCTGCTGGTGCAG 1702
1114 GluGluGlnThrSerLysIleLeuMetGlnTyrGlnAlaArgLeuGlu 1130
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1703 GAGCAGCAGATGCAGAAGCTGCTGGAATACAAGGCCCGACTGGAGGA 1752
1130 nSerGluLysArgLeuArgGlnGlnGlnValGluGluLysAspSerGlnIleL 1147
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1753 CAGCAGGAGCGGCTCCGAAGACAGCAGGAAGAAAAGATAGCCAGATGA 1802
1147 ySerIleIleGlyArgLeuMetLeuValGluGluLeuArgArgAsp 1163
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1803 AAAGCATCATCAGCAGGCTAATGGCTGTGGAAGAGGAACCTGAAGAAGGAT 1852
1164 HisProAlaMetAlaGluProLeuProGluProLysLysArgLeuLeuAs 1180
||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
1853 CATGCTGAGATG...CAAGCAGTATTGATGCAAAAGCAGAAAATAATTGA 1899
1180 pAlaGln.....Arg..Glyse 1185
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1900 TGCACAGGTCATAAATGGAGATGAGATTATTCAAACAGGAAAAACGGATC 1949
1185 rPheProProTyrValGlnGlnThrArg.....ValArgTrpPro. 1198
| |||||:|||| ||| :||| :||| :||| :||| :||| :||| :|||
1950 GTGT...CCCTGGATTACGCCAACACCACTGATGAGCGCGCTGACCCA 1996
1199 .....HisLeuGlyThrAlaTrp.....ProProGln 1207
:: :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1997 AGTGAAGGAGCGGTACAGCATGCAGGTCCGCAATGGCATCTCCCCCACC 2046
1208 ProHisProProHisProGlyCysArg.....SerGlnArgThrAl 1221
||| ||||:|||| ||| :||| :||| :||| :||| :||| :||| :|||
2047 ACCCCACCAAGCTTCCATCACGGAGAATGGTGAATTCAAAAACAGCAGC 2096
1221 aSerSerGlyThr.....ProGlnT 1228
|:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2097 TGCTGACGGGCTTTGTCTGTGGAGGAGACAGAAGAAATTGACCCACTC 2146
1228 hrThrSerProProSerIleThrAspLeuLeuProCysAlaProTyrPro 1244
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2147 TCCTATCTCCA.....GACCTT.....TACCTA 2169
1245 GlyProProSerValThrAspLeuLeuProSer 1255
:::|||||:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2170 GCCCCTCCAGGTTTACAGAAATGTTGCTACTTCA 2202

```





OM of: US-09-294-298-6 to: EST:\* out\_format : pfs

Date: Jan 18, 2001 9:24 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool/US09294298/runat\_17012001\_125501\_27745/app\_query.fasta\_1.4359  
-DB=EST -QFMT=fastap -SUFFIX=first -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=2000000000 -USER=US09294298 @CGN1\_1\_7119 -NCPU=6  
-ICPU=3 -LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-294-298-6

Query length: 1325

Database: EST:\*

Database sequences: 7991742

Database length: -791223438

Search time (sec): 4688.750000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_est37:AV590698	+ 933.00	1066.68	2.1e-50	571	! AV590698 AV590698 Bos taurus brain
gb_gss21:AZ399131	- 764.00	874.89	1.0e-39	474	! AZ399131 LM0164P14R Mouse 10kb
gb_est44:AW779747	- 737.50	841.95	6.9e-38	593	! AW779747 hn85f10.x1 NCI_CGAP_K1
gb_est44:AW421227	- 731.50	834.99	1.7e-37	597	! AW421227 fj94f10.x1 zebrafish g
gb_est71:BE490923	- 694.50	791.97	4.2e-35	628	! BE490923 db38g08.x1 Blackshear/
gb_est22:AI606711	+ 660.00	753.57	5.8e-33	566	! AI606711 ml58d01.y1 Stratagene
gb_est71:BE463433	- 628.50	719.70	4.5e-31	463	! BE463433 hw23d06.x1 NCI_CGAP_K1
gb_est22:AI577567	- 600.00	685.89	3.4e-29	511	! AI577567 UI-R-YO-vk-d-04-0-UI.s
gb_est41:AW205989	- 571.50	654.04	2.0e-27	474	! AW205989 UI-H-BII-afy-d-12-0-UI
gb_est28:AJ396547	+ 566.50	643.77	7.6e-27	707	! AJ396547 AJ396547 dkfz426 Gallu
gb_est11:AA709010	- 563.00	645.47	6.1e-27	427	! AA709010 zf94h11.s1 Soares_Pine
gb_est22:AI609604	- 548.50	627.99	5.7e-26	460	! AI609604 tn91e09.x1 NCI_CGAP_HN
gb_est9:AA554056	- 543.50	622.53	1.2e-25	449	! AA554056 n101g07.s1 NCI_CGAP_Cc
gb_est27:AI958829	+ 537.50	613.24	3.8e-25	555	! AI958829 fd22g01.y1 zebrafish w
gb_est40:AW136165	- 531.50	608.91	6.6e-25	443	! AW136165 UI-H-BII-abz-d-09-0-UI
gb_est10:AA691827	+ 517.50	592.62	5.3e-24	452	! AA691827 vt05d01.r1 Barstead m
gb_gss26:CNS02VGV	+ 501.00	564.47	2.0e-22	1020	! AL215753 Tetraodon nigroviridis
gb_est74:BE695645	- 486.00	554.98	6.7e-22	516	! BE695645 MR1-BT0800-030700-002-
gb_est9:AA575948	- 484.50	556.76	5.3e-22	379	! AA575948 nm56e05.s1 NCI_CGAP_BH
gb_est22:AI562171	- 484.50	556.67	5.4e-22	382	! AI562171 vw73f04.x1 Stratagene
gb_est41:AW204612	- 481.00	553.74	7.8e-22	347	! AW204612 UI-H-BII-aej-c-02-0-UI
gb_est2:AA107246	+ 480.00	549.90	1.3e-21	440	! AA107246 ml58d01.r1 Stratagene
gb_est39:AW076911	+ 473.00	544.63	2.5e-21	345	! AW076911 fj03d08.y1 zebrafish a
gb_est76:BE848541	+ 464.00	528.11	2.1e-20	596	! BE848541 uw39f05.y1 Soares_thy
gb_gss26:CNS02JDV	+ 461.50	519.62	6.2e-20	979	! AL200092 Tetraodon nigroviridis
gb_gss26:CNS02JUG	- 457.50	516.87	8.9e-20	832	! AL200329 Tetraodon nigroviridis
gb_gss14:AQ993964	+ 454.00	515.11	1.1e-19	682	! AQ993964 RPCI-23-241C10.TV RPCI
gb_gss26:CNS01XQ5	+ 452.50	514.27	1.2e-19	631	! AL172022 Tetraodon nigroviridis
gb_gss25:CNS00HWF	- 425.50	478.01	1.3e-17	1006	! AL073826 Drosophila melanogast
gb_gss26:CNS03ANH	+ 419.50	471.48	3.0e-17	975	! AL235430 Tetraodon nigroviridis
gb_est45:AW527364	- 418.50	483.06	6.8e-18	317	! AW527364 UI-R-B01-ajj-e-01-0-UI
gb_est59:BB321064	+ 417.50	480.38	9.6e-18	363	! BB321064 BB321064 RIKEN full-1e
gb_est26:AI874961	- 416.50	480.11	9.9e-18	336	! AI874961 ul27f03.x1 Sugano mous
gb_gss27:CNS04DL0	+ 416.50	469.24	4.0e-17	877	! AL285985 Tetraodon nigroviridis
gb_est4:AA258366	- 401.50	463.49	8.3e-17	319	! AA258366 zr59e01.s1 Soares_Nhm
gb_gss27:CNS041WW	+ 401.00	450.84	4.2e-16	926	! AL270761 Tetraodon nigroviridis
gb_est47:AW656799	+ 394.00	448.87	5.4e-16	542	! AW656799 109245 MARC lBOV Bos t
gb_est61:BB394547	+ 383.50	442.62	1.2e-15	325	! BB394547 BB394547 RIKEN full-1e
gb_est23:AI650331	+ 379.00	438.52	2.1e-15	296	! AI650331 wa18f01.x1 NCI_CGAP_K1
gb_est14:AA953060	- 368.50	426.63	9.4e-15	292	! AA953060 oo01g09.s1 Soares_NFL

gb\_est59:BB312776 + 365.50 422.54 1.6e-14 309 ! BB312776 BB312776 RIKEN full  
gb\_est75:BE773015 - 365.50 417.53 3.0e-14 481 ! BE773015 RCI-FT0134-170700-0  
gb\_est19:AI327335 - 358.00 405.50 1.4e-13 651 ! AI327335 mp74c11.x1 Soares\_t  
gb\_est66:BE085114 - 357.50 412.45 5.8e-14 335 ! BE085114 CM2-BT0664-240200-1  
seq\_name: gb\_est37:AV590698

seq\_documentation\_block:

LOCUS AV590698 571 bp mRNA EST 29-AUG-2000  
DEFINITION AV590698 Bos taurus brain fetus Bos taurus cDNA clone ElBR014B04  
5', mRNA sequence.

ACCESSION AV590698

VERSION AV590698.1 GI:9701691

KEYWORDS EST.

SOURCE cow.

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 571)

AUTHORS Sugimoto,Y., Hirotsune,S., Takasuga,A., Itoh,R., Jitohzono,A. and  
Suzuki,H.

TITLE bovine cDNA sequencing

JOURNAL Unpublished (2000)

COMMENT Contact: Yoshikazu Sugimoto

Animal Genetics Division

Shirakawa Institute of Animal Genetics

Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan

Tel: 81-248-25-5641

Fax: 81-248-25-5725

Email: kazusugi@cocoa.ocn.ne.jp

Single pass sequencing.

This clone was obtained from a polyA-deleted cDNA library.

Location/Qualifiers

FEATURES

source 1..571

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone="ElBR014B04"

/clone\_lib="Bos taurus brain fetus"

/tissue\_type="brain"

/dev\_stage="fetus"

/lab\_host="DH10B"

/note="Vector: pZL1; Site\_1: SalI; Site\_2: NotI; Poly A

was deleted from a NotI site"

BASE COUNT 124 a 161 c 169 g 116 t 1 others

ORIGIN

alignment\_scores:

Quality: 933.00 Length: 186

Ratio: 5.043 Gaps: 1

Percent Similarity: 99.462 Percent Identity: 99.462

alignment\_block:

US-09-294-298-6 x AV590698 ..

Align seg 1/1 to: AV590698 from: 1 to: 571

331 lysGluGluValAlaSerAlaLeuValHisIleLeuGlnSerThrGlyLy 347

|||||

13 AAGGAGGAGTCCGTCAGTCTCTGTTCACATCTCGAGGTGACGAGCA 62

347 salaLysAspPheLeuSerAspMetAlaMetSerGluValAspArgPhe 364

|||||

63 GGCCAAGGACTTCCTTCAGACATGGCCATGTCGAGGTGACCGTTCA 112

364 etGluArgGluHisLeuIlePheArgGluAsnThrLeuAlaThrLysAla 380

|||||

113 TGGAACGAGAACACCTCATATTCCGCGAGAACACGCTCGCACTAAAGCC 162

381 ileGluGluTyrMetArgLeuIleGlyGlnLysTyrLeuLysAspAla 397

|||||

163 ATAGAAGAGTACATGAGACTGATTGGTTCAGAAATACCTCAAGGATGCCAT 212

```
397 eGlyGluPheIleArgAlaLeuTyrGluSerGluAsnCysGluValA 414
|||||
213 CGGGGAATTCATCCGTCGTATGATCTGAGGAGAACTGTGAGGTGG 262
414 spProIleLysCysThrAlaSerSerLeuAlaGluHisGlnAlaAsnLeu 430
|||||
263 ACCCCATCAAGTGCACGGCATCCAGTTTGGCTGAGCACCAGGCCAACCTG 312
431 ArgMetCysCysGluLeuAlaLeuCysLysValValAsnSerHisCysVa 447
|||||
313 CGGATGCTGTGAGTTGGCCCTGTGCAAGGTGGTCAACTCCCATTCGCT 362
447 lPheProArgGluLeuLysGluValPheAlaSerTrpArgLeuArgCysA 464
|||||
363 GTTCCCGAGGGAGCTGAAGAGGTGTTGCGTCTCGGCGACTGCGCTGTG 412
464 laGluArgGlyArgGluAspIleAlaAspArgLeuIleSerAlaSerLeu 480
|||||
413 CAGAGCGGGCGCGAGAGGACATTCGGACCGGCTGATCAGCGCCTCGCTC 462
481 PheLeuArgPheLeuLeuCysProAlaIleMetSerProSerLeuPhe.GlyL 497
|||||
463 TTCCTGCGCTTCTCTGCGCGCCATTATGTGCGCCAGCCTNTTCGGGGC 512
497 euMetGlnGluTyrProAspGluGlnThrSerArgThrLeuThrLeuIle 513
|||||
513 TCATGCAGGAGTACCCAGATGAGCAAACCTTCGAGAACCCCTCACGCTCATC 562
514 AlaLys 515
|||||
563 GCCAAG 568
```

seq\_name: gb\_gss21:AZ399131

seq\_documentation\_block: 474 bp DNA GSS 03-OCT-2000

LOCUS AZ399131 1M0164P14R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
DEFINITION clone UUGC1M0164P14 R, DNA sequence.

ACCESSION AZ399131

VERSION AZ399131.1 GI:10514203

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 474)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

CONTACT: Robert B. Weiss

UNIVERSITY OF UTAH Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0164 row: P column: 14

Seq primer: CACACAGGAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 474.

Location/Qualifiers

1..474

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0164P14"

FEATURES

source

/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptored DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptored mouse DNA was annealed to  
adaptored vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 92 a 151 c 131 g 100 t  
ORIGIN

alignment\_scores:

Quality: 764.00 Length: 142

Ratio: 5.380 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.296

alignment\_block:

US-09-294-298-6 x AZ399131/rev ..

Align seg 1/1 to reverse of: AZ399131 from: 1 to: 474

141 LysAspAsnSerArgArgValAspAsnValLeuLysLeuTrpIleG1 157

:::|||||

426 CAGGACACAGCCCGCGAGTAGATAACGTGCTGAGAGCTATGGATCATAGA 377

157 uAlaArgGluLeuProProLysLysArgTyrTyrCysGluLeuCysLeuA 174

|||||

376 GGCTCGAGAGCTGCCCCCAAGAGAGAGATATTACTGTGAGCTGTGCCTGG 327

174 sPaspMetLeuTyrAlaArgThrThrSerLysProArgSerAlaSerGly 190

|||||

326 ACGACATGCTGTATGCACGACACACCTCCAAGCCCCGCTCGGCTTCAGGA 277

191 AspThrValPheTrpGlyGluHisPheGluPheAsnAsnLeuProAlaVa 207

|||||

276 GACACCGCTTTTGGGGCGGAGCATTGAGTTTACACACCTGCCTGCCGT 227

207 lArgAlaLeuArgLeuHisLeuTyrArgAspSerAspLysLysArgLysL 224

|||||

226 CCGGGCCCTTCGGCTGCATCTGTACCGTGACTCAGACAAAAGCGGAAGA 177

224 ysAspLysAlaGlyTyrValGlyLeuValThrValProValAlaThrLeu 240

|||||

176 AGGACAAGCTGGCTACGTTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 127

241 AlaGlyArgHisPheThrGluGlnTrpTyrProValThrLeuProThrG1 257

|||||

126 GCTGGGGCCCACTTCACAGAGCAGTGGTACCCCGTGACCTTGGCCGACAGG 77

257 ySerGlyGlySerGlyGlyMetGlySerGlyGlyGlyGlyGlySerGlyG 274

|||||

76 CAGTGGGGGCTCTGGGGGCGATGGGCTCGGGGGGAGGAGGGTTCAGGGG 27

274 lyGlySerGlyGlyLysGlyLysGly 282

|||||

26 GTGGCTCAGGGGGCAAGAGGGAAGGA 1

seq\_name: gb\_est49:AW779747





Email: zbrafish@watson.wustl.edu  
 cDNA Library Preparation: Leonard Ira Zon DNA Sequencing by:  
 Washington University Genome Sequencing Center Clone Distribution:  
 Genome Systems, St. Louis, Missouri (web address:  
 www.genomesystems.com) (email contact: info@genomesystems.com) and  
 Research Genetics, Huntsville, Alabama (web address: www.resgen.com  
 ) (email contact: info@resgen.com) and  
 Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:  
 www.rzpd.de)

Possible reversed clone: polyT not found  
 Seq primer: T7 ET from Amersham  
 High quality sequence stop: 409.

## FEATURES

source

```
1. .597
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="zebrafish gridded kidney"
/sex="mixed"
/tissue_type="kidney pooled from 300 wild type adults"
/lab_host="XLOR"
/Note="Organ: kidney; Vector: pBK-CMV; Site_1: EcoRI;
Site_2: XhoI; Oligo dt cDNA library constructed from mRNA
pooled from pooled kidney tissue from 300 adult
zebrafish."
```

BASE COUNT 132 a 153 c 150 g 162 t  
 ORIGIN

alignment\_scores: Quality: 731.50 Length: 202  
 Ratio: 4.041 Gaps: 2  
 Percent Similarity: 89.604 Percent Identity: 67.327

## alignment\_block:

US-09-294-298-6 x AW421227/rev ..

Align seg 1/1 to reverse of: AW421227 from: 1 to: 597

```
18 GlyPheLeuSerArgArgLeuLysSerSerIleLysArgThrLysSerGI 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
591 GGTTTTAAAGTAAACGACTGATAGGATCCATCAAAAGGACGAAGCCCA 542

34 nProLysLeuAspArgThrSerSerPheArgGlnIleLeuProArgPhea 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
541 GACCAAGCTGGACCGCAACACACAAAGTTTCAGA.....CTCCCATCGCTGC 498

51 rgSerAlaAspHisAspArgAlaArgLeuMetGlnSerPheLysGluser 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
497 GACCAACAGAAATGACAGGTGACAGGCTGACAGGCTGCCCAATGAAGGAGTCC 448

68 HisSerHisGluSerLeuLeuSerProSerSerAlaAlaGluAlaLeuGI 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
447 TGTTCCTCATGAGTCTTTGCTGAGTCTGCGAGCGCTGTGAAGCTCTGGA 398

84 uLeuAsnLeuAspGluAspSerIleIleLysProValHisSerSerIleL 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
397 TCTGAGTATGAGGAGGAGTGTCTACATCAAAACCTTTACACAGCAGCATCC 348

101 euGlyGlnGluPheCysPheGluValThrThrSerSerGlyThrLysCys 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
347 TAGGACAAGATTCTGCTTTGAGGTGGCGTACTCAGGTGGAAGTAAGTGC 298

118 PheAlaCysArgSerAlaAlaGluArgAspLysTrpIleGluAsnLeuGI 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
297 TTCAGCTGTCTCTGCTGCTGAACGGGACAAATGGATGGAGAACCTCAA 248

134 nArgAlaValLysProAsnLysAspAsnSerArgArgValAspAsnVall 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
247 AAGAACTGTGCAACCTTAATAGGACAAATGCCGGGGGGGGAACCTGCC 198

151 euLysLeuTrpIleIleGluAlaArgGluLeuProProLysLysArgTyr 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
197 TCCGCCTGTGGATCATTTGAGGCAAAAGACCTGCCACCAAAAGAAATAT 148
```

```
168 TyrCysGluLeuCysLeuAspAspMetLeuTyrAlaArgThrSerLy 184
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
147 TTCTGTGAGCTGTCTGTGATGACATGCTGTACGCTCGTACCACGACAA 98

184 sProArgSerAlaSerGlyAspThrValPheTrpGlyGluHisPheGluP 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
97 AACCCGCTCC.....GACTGCCTGTCTTGGGGGGAACACTTGTGAGT 57

201 heAsnAsnLeuProAlaValArgAlaLeuArgLeuHisLeuTyrArgAsp 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
56 TTTCCGGCTCCCGCTCCATGAAGAGCATCACCATACATCTACCCGGGAT 7

218 SerAsp 219
|||
6 GTGGAC 1

seq_name: gb_est71:BE490923

seq_documentation_block: 628 bp mRNA EST 31-JUL-2000
LOCUS BE490923 db38g08.x1 Blackshear/Soares normalized xenopus egg library xenopus
DEFINITION laevis cDNA clone IMAGE:3300830 3' similar to TR:095174 095174
NGAP.; mRNA sequence.
ACCESSION BE490923
VERSION BE490923.1 GI:9610456
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodidae; Xenopus.
1 (bases 1 to 628)
REFERENCE Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
AUTHORS B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
WashU Xenopus EST project, 1999
Unpublished (1999)
Other_ESTs: db38g08.y1
Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Bento Soares and M. Fatima Bonaldo
(University of Iowa). DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 454.
Location/Qualifiers
1. .628
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3300830"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/Note="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
PolyA-selected mRNA was prepared from unfertilized xenopus
laevis eggs. The library was constructed in the vector
pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
```



```

1706 GTTCATGATAGTCTCTTTGGCAGGCCAAACTGTTTTCAGATAGTACTCA 1755
112 rSerGlyThrLys.....CysPheAlaCysArgSerAlaA 124
: : : : : ||||| : : : : :
1756 GCATTTAGTGAAGAACAATACATCTTTTACTTTGCAGGAGAAACTCCAG 1805
124 laGluArgAspLysTrpIleGluAsnLeuGln.....Arg 135
: : : : : |||: : : : : ||||| : : : : :
1806 AACAAGCAGAGGATTGGATGAAGGCTCTGCAGGSCATTTTGCAATTACGG 1855
136 AlaValLysProAsnLysAspAsnSerArg...ArgValAspAsnValLe 151
: : : : : : : : : : : ||||| : : : : :
1856 AAAAGTAGTCCAGGGACATCCAATAAACGCCCTTCGTGAGGTTCAGCAGCCT 1905
151 uLysLeuTrpIleIleGluAlaArgGluLeuLeuProLys.....LysA 166
- ||| ||| |||||: : : : : ||||| ||| : : : : :
1906 TGTTTTACATATTGAAGAAGCCCATAAACTCCAGTAAACATTTTACTA 1955
166 rgTyrTyrCysGluLeuCysLysLeuAspMetLeuTyrAlaArgThrThr 182
: : |||||: : : : : |||: : : : : |||: : : : :
1956 ATCCATATTGTAACATCTACCTGAATAGTGTCGAAGTAGCAAAACTCAT 2005
183 SerLysProArgSerAlaSerGlyAspThrValPheTrpGlyGluHisph 199
: : : : : : : |||: : : : : |||: : : : :
2006 GCAAGG.....GAAGGGCAAAACCCAGTATGGTCAGAAGAGTT 2043
199 eGluPheAsnAsnLeuProAla...ValArgAlaLeuArgLeuHisLeuT 215
- |||: : : : : ||||| : : : : : : : : : : |||
2044 TGTCTTTGATGATCTTCCTCTGACATCAATAGATTTGAAATAACTCTT. 2092
215 yrArgAspSerAspLysLysArgLysLysAspLysAlaGlyTyrValGly 231
: : : : : ||| |||||: : : : : ||| : : : : :
2093 .....AGTATAAAACAAGAAAGCAAGATCCTGATATCTTTA 2131
232 LeuValThrValProValAlaThrLeuAlaGlyArgHisPheThrGluG1 248
: : : : : : : : : : ||| ||| |||: : : : :
2132 TTTATGCGCTGCCAGTTGAGCCGATTACAGAAAGGCGCATGCCACAGATGA 2181
248 nTrpTyrProValThr.....LeuProThrGlySerGlyGlySerGlyG 263
: |||: : : : : : : |||
2182 ATGTTTCTGCTCAGCTCCCATATACCATTA..... 2212
263 lyMetGlySerGlyGlyGlyGlySerGlyGlySerGlyGlyLys 279
2212 ..... 2212
280 GlyLysGlyGlyCysPro...AlaValArgLeuLysAlaArgTyrGlnTh 295
||| ||| : : : : : : : : : : |||||: : : : :
2213 ...AAAGGTATTGAACCGAGTCCCTGCGTGTGAGCAGCATACTCTAT 2259
295 rMetSerIleLeuProMetGluLeuTyrLysGluPheAlaGluTyrValT 312
: : : : : ||| ||| |||: : : : : ||| : : : : :
2260 GGAAAAAATCATGCCAGAGAAGAGTACAGTGAATTTAAAGAGCTTATAC 2309
312 hrAsnHisTyrArgMetLeuCysAlaValLeuGluProAlaLeuAsnVal 328
: : : : : : : : : : : |||: : : : :
2310 TGCAAAAGGAACCTTCATGTAGTCTATGCTTTATCA.....CATGTA 2350
329 LysGlyLysGluGlu.....ValAlaSerAlaLeuValHisIleLeuG1 343
|||: : : : : : : ||||| |||: : : : : |||: : : : :
2351 TGTGGACAAGACCGAACACTACTGGCCAGCATCCTACTGAGGATTTTCT 2400
343 nSerThrGlyLysAlaLysAspPheLeuSerAspMetAlaMetSerGluV 360
: : : : : : : ||| : : : : : |||: : : : :
2401 TCACGAAAAGCTTGAATCGTTGTTGTTATGCACACTAAATGACAGAGAAA 2450
360 alaPArgPheMetGluArgGlu...HisLeuIlePheArgGluAsnThr 375
: : : : : ||||| ||| : : : : : |||: : : : :
2451 TAAGC.....ATGGAAGATGAAGCCACTACCTATTTCGAGCCACAAACA 2494
376 LeuAlaThrLysAlaIleGluGluTyrMetArgLeuIleGlyGlnLysty 392
|||||: : : : : |||: : : : : |||: : : : :
2495 CTTGCAAGCACCTTGTATGAGCAGAGTATATGAAGCCACTGCTACACAGTT 2544

```

392 rLeuLysAspAlaIleGlyGluPheIleArgAlaLeuTyrGluSerGluG 409  
 : : : : | | | : : : : | | : : : | | | | : : :  
 2545 TGTTTCATCATGCTTTGAAAGACTCTATTTTAAAGATAAATGGAAGCAAGC 2594  
 409 luAsnCysGluValAspProIleLysCysThrAlaSerSerLeuAlaGlu 425  
 : : : : | | | | : : : : | | | : : : : : : : : : :  
 2595 AGTCTTGTCAGTTAAGTCCATCAAAAGTTAGAAAAAATGAAGATGTAAC 2644  
 426 HisGlnAlaAsn.....LeuArgMetCysCysGluLeuAlaLeuCysLy 440  
 :  
 2645 ACTAATTTAAACACACCTATTGAACATACCTTTTCAGAGCTT...GTGGAGAA 2691  
 440 sValValAsnSerHisCysValPheProArgGluLeuLysGluValPheA 457  
 | : : : : : : : : : : | | | : : : : : : : : : :  
 2692 AATATTTCATGGCTTCAGAAATACTTCCACCGACATTCAGATATATTATG 2741  
 457 laSerTrpArgLeuArgCysAlaGluArgGly..... 467  
 : : | | | : : : : : : : : : : : : : : : : : :  
 2742 GG.....TGTTACAGAAATCTGTTTCAGCATAAAGTGGCCT 2776  
 468 ...ArgGluAspIleAlaAspArgIleuIleSerAlaSerLeuPheLeuAr 483  
 : : : : | | | : : : : : : : : : : : : : : : : :  
 2777 ACAAAATACCAACCATGAGAACAAAGAGTTGTTAGTGGTTTGTCTTCG 2826  
 483 gPheLeuCysProAlaIleMetSerProSerLeuPheGlyLeuMetGlnG 500  
 | : : : : | | | | | | | : : : : : : : : : : : : : : :  
 2827 ACTCATCTGCTCCTGCCATCCTGAATCCACGGATGTTCAATATCATCTCAG 2876  
 500 luTyrProAspGluGlnThrSerArgThrLeuThrLeuIleAlaLysVal 516  
 : : | | | : : : : : | | | | | | | : : : : : | | |  
 2877 ATTCCTCCATCTCCTATTGCTGCAAGAACACTGATATTAGTGGCTAAATCT 2926  
 517 IleGlnAsnLeuAlaAsnPheSerLysPheThrSerLysGluAspPheLe 533  
 : : | | | | | | | | : : : : | | : : | | | | : : : :  
 2927 GTGCAGAACTTAGCAAAATCTGTGGAATTTGGAGCTAAGGAGCCCTACAT 2976  
 533 uGlyPheMetAsnGluPheLeuGluLeuGluTyrGlySerMetGlnGlnP 550  
 : : : : | | | : : : : : : : : : : | | | : : : :  
 2977 GGAAGGTGTCATCCATTTCATCAAAAGCAACAACATCGTATGATCATGT 3026  
 550 heLeuTyrGluIleSerAsnLeuAspThrLeuThrAsnSerSerSerPhe 566  
 | | | | | | : : : : : : : : | | : : : : : : : : : :  
 3027 TTTTAGATGAACCTGGGAATGTACCTGAACCTCCGGACACTACAGAG... 3073  
 567 GluGlyTyrIleAspLeuGlyArgGluLeuSerThrLeuHisAlaLeuLe 583  
 : : : : | | | : : : : : | | : : : : : | | | : : :  
 3074 CATCTAGAACGGACCTGTCCCGTGATTTACCACCATTCATGAGATTG 3123  
 583 u.....TrpGluValLeuProGlnLeuSerLysGlu..... 593  
 : : : | | | | | | : : : : : : : : : : : : : :  
 3124 CGTGGCTCATTCAGATGAACCTCGAAGCTCAGTAATGAGCGGTGGTGCAC 3173  
 594 .....AlaLeuLeuLysLeuGlyProLeuProArgLeuLeuSerAsp 607  
 : : | | | | | : : : : : : : : : : : : : : : : :  
 3174 AGCAGCACGTATTGAAAAAGCTTCTGGCTATACACAGAACCTGCTTCACAA 3223  
 608 IleSerThrAlaLeuArgAsnProAsn...IleGlnArgGlnProSerAr 623  
 : : : : : : : : | | | : : : : : : : : | | | | | |  
 3224 AAACAAACCCAGTATACAAAAACCAATGATGTCAAGGTAGCAGCCTTCGCC 3273  
 623 gGln 624  
 | | |  
 3274 CCAG 3277

seq\_name: /SIDS6/gc9data/geneseq/geneseq/NA1991.DAT:Q10988

seq\_documentation\_block:

ID Q10988 standard; cDNA; 4307 BP.

XX  
AC Q10988;

XX  
DT  
22-MAY-



OM of: US-09-294-298-6 to: Issued\_Patents\_NA:\* out\_format : pfs  
 Date: Jan 19, 2001 7:58 AM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
 -Q=/cgn2\_1/USPTO\_spool/US09294298/runat\_17012001\_125501\_27768/app\_query.fasta\_1.4359  
 -DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000  
 -USER=US09294298 @CGN1\_1\_163 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
 -WAIT -THREADS=1

Search information block:

Query: US-09-294-298-6  
 Query length: 1325  
 Database: Issued\_Patents\_NA:\*  
 Database sequences: 280836  
 Database length: 80580151  
 Search time (sec): 156.930000

score\_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-190-687B-24	..	437.50	420.83	6.9e-16	3456
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-190-687B-7	+	437.50	418.67	9.1e-16	4307
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-909-954-3	+	371.00	355.22	3.1e-12	3117
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-909-954-1	+	360.00	344.10	1.3e-11	3148
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-557-139-1	+	276.00	253.38	1.5e-06	6152
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-781-802-7	-	253.00	236.91	1.2e-05	3147
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-551-531-1	+	251.00	228.49	3.6e-05	6060
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-449-933-1	+	251.00	224.68	5.8e-05	8937
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-411-389-1	+	251.00	222.91	7.3e-05	10706
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-13	-	251.00	220.29	0.0001	13987
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-7	-	251.00	208.98	0.0004	44377
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-198-1	-	251.00	208.98	0.0004	44377
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-574-959A-6	+	248.50	230.30	2.8e-05	3901
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-945-283-1	-	247.50	221.74	8.4e-05	8438
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-149-097D-22	+	247.00	222.00	8.2e-05	7808
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-949-386-22	+	247.00	222.00	8.2e-05	7808
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-450-562-22	+	247.00	222.00	8.2e-05	7808
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-1	-	246.00	204.21	0.0008	43280
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-779-801-3	+	243.50	225.99	4.9e-05	3632
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-343-785-7	+	242.50	230.59	2.7e-05	2051
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-462-221-7	+	242.50	230.59	2.7e-05	2051
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-946-458-7	+	242.50	230.59	2.7e-05	2051
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-343-785-1	+	242.50	230.31	2.8e-05	2111
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-462-221-1	+	242.50	230.31	2.8e-05	2111
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-946-458-1	+	242.50	230.31	2.8e-05	2111
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-574-959A-8	+	242.00	225.69	5.1e-05	3211
/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCT-US95-14418-1	+	241.50	225.70	5.1e-05	3048
/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCT-US95-15327-1	+	241.50	225.70	5.1e-05	3048
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-779-801-5	+	241.50	224.08	6.3e-05	3596
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-945-283-1	+	240.50	214.72	0.0002	8438
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-261B-6	+	240.50	202.64	0.0010	28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-837-6	+	240.50	202.64	0.0010	28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-342-6	+	240.50	202.64	0.0010	28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-646A-6	+	240.50	202.64	0.0010	28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-076A-6	+	240.50	202.64	0.0010	28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-764-233A-4	+	240.50	202.64	0.0010	28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-335A-6	+	240.50	202.64	0.0010	28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-729-214-6	+	240.50	202.64	0.0010	28958
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-028-934-6	+	240.50	202.64	0.0010	28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-764-233A-1	+	240.50	197.41	0.0019	49377

/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-07-882-292-1 + 239.00 223.93 6.4e-05 2630  
 /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-331-644-1 + 239.00 223.93 6.4e-05 2630  
 /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:PCT-US93-04102-1 + 239.00 223.93 6.4e-05 28  
 /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-779-801-4 + 239.00 221.48 8.7e-05 3632  
 /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-804-227C-7 + 238.50 196.45 0.0022 44377

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-190-687B-24

seq\_documentation\_block:

; Sequence 24, Application US/08190687B  
 ; Patent No. 5760203  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wong, Gail L.  
 ; APPLICANT: Martin, George  
 ; APPLICANT: McCormick, Francis P.  
 ; APPLICANT: Rubinfeld, Bonnie  
 ; APPLICANT: O'Rourke, Edward C.  
 ; APPLICANT: Clark, Robin  
 ; TITLE OF INVENTION: GAP Gene Sequences  
 ; NUMBER OF SEQUENCES: 27  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/190,687B  
 ; FILING DATE: 02-FEB-1994  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/774,644  
 ; FILING DATE: 11-OCT-1991  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/260,807  
 ; FILING DATE: 21-OCT-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/230,761  
 ; FILING DATE: 10-AUG-1988  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Gass, David A.  
 ; REGISTRATION NUMBER: 38,153  
 ; REFERENCE/DOCKET NUMBER: 27527/31898  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 24:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3456 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 100..2709  
 ; US-08-190-687B-24

alignment\_scores:  
 Quality: 437.50 Length: 568  
 Ratio: 1.326 Gaps: 22  
 Percent Similarity: 58.099 Percent Identity: 26.232

alignment\_block:  
 US-09-294-298-6 x US-08-190-687B-24

seq\_documentation\_block: